

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2002, 12:29:46 : Search time 1930.26 Seconds
(without alignments)
8255.089 Million cell updates/sec

Title: US-09-923-304-1
Perfect score: 722
Sequence: 1 cgcctaacgcctccaacatg.....aaggctcttttgacctatt 722

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1737556 seqs, 10463268293 residues
- al number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pi.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred, No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query
No. Score Match Length DB ID Description

1	722	100.0	722	9	D87735
2	695.8	96.5	808	9	BC009294
3	594.4	96.2	721	9	BC019651
4	595	95.0	833	9	BC005134
5	584.8	94.8	760	9	BC000606
6	572.4	93.1	713	9	BC022805
7	568	92.5	753	9	HUMNAE
8	558.8	91.8	179523	2	AC026104
9	558.8	91.2	191461	2	AC078814
10	547.6	91.1	701	9	HS016738
11	547.6	89.7	175099	2	AC017029
12	535.8	82.5	632	9	HS02982
13	499.4	69.2	212421	2	AC109591
14	485.4	67.2	208632	2	AC073711
15	482.2	66.8	230760	2	AC073714
16	471.2	65.3	715	10	NR1B0L14
17	441.4	61.1	173493	9	AC093115
18	441.4	61.1	184447	9	AC012519
19	431.6	59.8	190162	2	AC097966
20	413.2	57.2	181238	2	AC094176
21	412.5	57.1	177490	2	AC095342
22	376.4	52.1	573	10	BC008177
23	352	48.8	7729	9	AB061822
24	338	46.8	321708	2	AC087142
25	334.8	46.8	338116	2	AC087159
26	315.4	43.7	176075	9	AC017079
27	312	43.2	185208	30	AC019204
28	312	43.2	187253	2	AC104186
29	302.4	41.9	166585	2	AC108039
30	302.4	41.6	178825	2	AC074251
31	302.4	41.6	178825	2	AC074251
32	302.4	41.6	178825	2	AC074251
33	263.8	36.5	406	5	AB046393
34	252.2	34.9	468	5	AF01569
35	248.4	34.4	457	5	TR0310911
36	219.4	30.4	1193	9	BC010079
37	182	25.2	281	11	HS0281P
38	169.2	23.4	638	3	RLA33201
39	151	20.9	167	11	G84294
40	147.8	20.5	283	6	A74662
41	147.8	20.5	283	6	A77641
42	146.4	20.3	184916	2	AC068998
43	146.4	20.3	226199	2	AC084069
44	129.6	18.0	152408	9	HS191318
45	120.2	16.5	163739	2	AC093718

ALIGNMENTS

RESULT	1
D87735	D87735
LOCUS	Homo sapiens mRNA for ribosomal protein L14, complete cds.
DEFINITION	D87735
ACCESSION	D87735
VERSION	D87735.1 GI:1620021
KEYWORDS	ribosomal protein L14,
SOURCE	Homo sapiens neonatal male umbilical cord vein endothelial cell
ORGANISM	Cell_line:HUE4 cdna to mRNA, clone_lib:PCMV-SPORT clone:thUE4-4.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	1 (bases 1 to 722)
REFERENCE	Mitsui.Y.
AUTHORS	Submitted (05-SEP-1996) Youji Mitsui, Agency of Industrial Science
TITLE	and Technology, National Institute of Bioscience and
JOURNAL	Human-Technology, Higashi 1-1, Tsukuba Science City, Ibaraki 305.
REFERENCE	Japan [E-mail:ttanaka@is.icc.u-tokai.ac.jp, Tel:+81-298-94-6070,
AUTHORS	Fax:+81-298-94-6095)
TITLE	2 (sites)
REFERENCE	Tanaka.M., Tanaka.T., Harata.M., Suzuki.T. and Mitsui.Y.
AUTHORS	triplet repeat-containing ribosomal protein L14 gene in
TITLE	

Qy		481	ctgtcgtgctgctgctgctgctgctgctgctgcaaaaaagatcacccgccgcga	540
Dd		481	CTGTCGTGCTGCTGCTGCTGCTGCTGCTAAAGTTCCAGCAAAAAGATCACCGGCCGA	540
Qy		541	gtaaaaggcttcacggcccagaagtcttcccgacaagaccagaccagaaagcagcgc	600
Dd		541	GTAANAAGGCTCAGGCCACAGAAGTTCCTGCCAGAANGCCCAGGCCAGANAGCGGC	600
Cy		601	ctgtcccaaagctcagaagggtccaataaaagactccagccccagaagcacctgtcccaagg	660
Dd		601	CTGTCCAAAAAGGTCAGAAGGGTCAAANAAGCTCAGGCCAGAAAGCACGTGCTCCAAGG	660
Qy		661	catttgccaagaagcataagttgaccaatcatataaaagtaataaagattcttttgacctg	720
Dd		661	CATTCGCAGANAAGCATNAAGTGCCATCATANAAGGTATAAAGGTTCTTTTTIGACTTG	720
Qy		721	tt 722	
Dd		721	TT 722	
RESULT 2				
BC009294				
LOCUS				
DEFINITION Homo sapiens, clone MGC:16644 IMAGE:4123062, mRNA, complete cd				
ACCESSION BC009294				
VERSION BC009294.1 GI:14242451				
KEYWORDS MGC.				
SOURCE human.				
ORGANISM Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleosteo-				
Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.				
REFERENCE				
AUTHORS Strausberg,R.				
TITLE Direct Submission				
JOURNAL Submitted (12-JUN-2001) National Institutes of Health, Mammali				
Gene Collection (MGC), Cancer Genomics Office, National Cancer				
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-259				
USA				
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT Contact: MGC help desk				
Email: cgapbs@mail.nih.gov				
Tissue Procurement: AICC				
CDNA Library Preparation: Rubin Laboratory				
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)				
Sequencing Center (NISC).				
Gaithersburg, Maryland:				
Web site: http://www.nisc.nih.gov/				
Contact: nisc.mgc@nih.gov				
Shenchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,				
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brook,				
Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legasp				
Lim,M., Maduro,O.L., Masilelo,C., Mastrian,S.D., McCloskey,J.C				
McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J.,				
Wongson,E.E., Touchman,J.W., Tsurgoon,C., Voigt,J.L., Walker,M.H.				
Zhang,L.-H. and Green,E.D.				
Clone distribution: MGC clone distribution information can be				
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.g				
Series: IRAL File: 26 Row: e Column: 3				
This clone was selected for full length sequencing because it				
passed the following selection criteria: Hexamer frequency ORF				
analysis, Similarity but not identity to protein.				
Location/Qualifiers				
1..808				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/clone="MGC:16644 IMAGE:4123062"				
/tissue_type="MH_Muscle, rhabdomysarcoma"				
/clone_lib="NIH_MGC_17"				
/lab_host="DH10B-R"				
/note="vector: pOT7"				
FEATURES				
source				

LOCUS BC019651 721 bp mRNA linear PRI 22-JAN-2000
DEFINITION Homo sapiens, ribosomal protein L14, clone MGC:25011 IMAGE:4454086.
ACCESSION BC019651
VERSION BC019651.1 GI:18044606
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 721)
AUTHORS Strausberg, R.
JOURNAL Direct Submission
Submitted (19-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hqsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gonaratne, P. H., Garcia, A. M., Lu, X., Hulyk, S. W., Hale, S. M.,
Yoon, V. S., Kowis, C. R., Lawrence, S., Martin, R. G., Muzny, D. M.,
Richards, S., Gibbs, R. A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 29 Row: b Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency CRF
analysis.
FEATURES
source Location/Qualifiers
1..721
/organism="Homo sapiens"
/db_xref="LocusID:9045"
/db_xref="taxon:9606"
/clone="MGC:25011 IMAGE:4454086"
/tissue_type="Kidney, hypernephroma"
/clone_lib="NIH_MGC_89"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
8..667
/codon_start=1
/product="ribosomal protein L14"
/protein_id="AAH19651.1"
/db_xref="GI:18044607"
/translation="MVFRFEVGRVAVSVFGPHAGKLVAIVDIDONRALVDFGCTOT
VRQAPFKCMQLTDFILKPFHSQVVRQAWOKADINTKWAIRWAKKI EAPERKRAK
KMTDFREKVMYKAKMWRNRI IKNEVKKKLOKAAALLKSPKAPGT KGTAAAAA
AAAKVPAKKITATSKRAPAQKVPAQKATGOKAPAPKAQKQKAPAKAFAPRASGKKK
A
BASE COUNT 228 a 167 c 179 g 147 t
ORIGIN
Query Match 96.2% Score 694.4; DB 9; Length 721;
Best Local Similarity 99.4%; Pred. No. 3.le-139;
Matches 708; Conservative 0; Mismatches 1; Indels 3; Gaps 1;
QY 11 tggcaacatggtgttcaggcgcttcgtggaggttggcgggatgacctatctcttgg 70
Db 1 TGCCAACATGGTGTTCAGGCGCTTCGTGGAGGTGGCGGGTTCCTATGCTCTCCTTGG 50
QY 71 acctcatgccggaataatggtcgcgattgtatgattattatcatcagaacagagcttgg 130
Db 51 ACCTCATGCCGGAATAATGGTTCGGATGTAGATGTTATGATCATCAGACAGGGCTTGGT 120


```

507 caaagctcagaaggtcaaaagctccagcccaagcccaagccacctctcaaggcatctg 566
|||||
605 CAAAAGCTCAGAAAGGTCAGAAAGTCCAGCCGACAGAAAGCAGCTGCTCAAAAGCAICTG 564
|||||
667 gcaagaagcatalaagtcgaatcatalaataaagtaataaaggtctcttaccatatt 722
|||||
665 GCAGAAAGCATAGTGGCAATCATAAAGTAATTAAGGTTCTTTTACCTTIGTT 720

RESULT 5
LOCUS BC000606 760 bp mRNA linear FRI 12-JUL-2001
DEFINITION Homo sapiens, Similar to ribosomal protein L1b, clone MGC:1544
IMAGE:3345054, mRNA, complete cds.
ACCESSION BC000606
VERSION BC000606.1 GI:12653648
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 760)
Direct Submission
Strausberg, R.
Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/PCID/DIP
CDNA Library Preparation: Pubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Einkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Masiello, C., Mastrian, S., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stenrups, S., Thomas, P.J.,
Tiongson, E., Touchman, J.W., Tsurganov, C., Venter, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IMAGE Plate: 5 Row: 9 Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: matched RNA gi: 4505600.
Location/Qualifiers
1. 760
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:1544 IMAGE:3345054"
/tissue_type="Skin, melanotic melanoma"
/clone_lib="NIH_MGC_20"
/lab_host="DH10B-P"
/notes="vector: pOTB7"
45. 695
/codon_start=1
/product="Similar to ribosomal protein L1b"
/protein_id="AAH00506.1"
/db_xref="GI:12653648"
/translation="MVFRFRFEVGRVAYVSPGPHAGFLVAIVGVIDONRALVDGFCIQ
VPQAMPKCMQLDFFILKFPFSAHQVYVPOAWOKA.LINIVKATIPRAKKIEAFEPKA
KMTDFDKVWKAKMKNRIKNEVKLQKAALEKAFKESDCKGTAAANAAAAA
KVPKATTAASKAPAKQVPAQKAPAKALPAPKALPAPKAPAKAPAKASGRKA"

BASE COUNT 241 a 179 g 188 g 152 t
ORIGIN

```

```

Query Match 94.8%; Score 684.8; DB 9; Length 760;
Best Local Similarity 98.1%; Pred. No. 3.6e-137;
Matches 708; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 caccataacccctcccaacatggtgttcagcgcgtcttcagaggttgagccgggtgacctatg 60
DB 28 CGCCCTAACCGCCGCCAACATGGTGTTCAGCGCGCTTCGTGGAGGTTGGCCGGTGGCCCTAIG 8~
QY 51 tctcctttgacacctatgccggaataattggtccgcatgttagatgttatgtgacaaaca 120
DB 88 TCTCCTTTGGACCTCAIGCGCGGNAATTTGGTCGGATTGTAGATGTTATTTGATCAGAACA 117
QY 121 gggctttggtcgaatgcaccttcacactcaagttagggagacagggccatgcctttcaagtga 130
DB 148 GGCGTTTGGTCGATGGACCTTGCACCTCAAGTAGGAGAGACAGGCCATTCCTTTTCAAGTGCA 207
QY 181 tcaagctcaatgatttcacctcaaatcttcacacatgccacacagaagaatattccgac 240
DB 208 TGCAGCTCACTGATTTTCATCTCAAGTTTCCGCACAGTGGCCACCAAGATATGTCGGAC 267
QY 241 aagcctggcagaagggcagacataatcacaaaatgggcagccacacgatgggccaagaaga 300
DB 258 AAGCCTGGCAGAGGCGAGACATCAATACAAAATGGCGACCCACACGATGGCCCAAGAAGA 327
QY 301 ttgaagccagagaagaagaagcccaagatgacagatbttgatcttttaagtataaagg 360
DB 328 TTGAGCCAGAGAGAAAGCAAGCCCAAGATGACAGATTTTATCGTTTATAAGTTATGAAGG 387
QY 361 caaagaaatgaggaacacacataatcaagaataaacttaagaadcttcaaaagcgactc 420
DB 388 CAAAGAAATGAGGAACAGAAATTAATCAGAAATGAAGTTAAGAAGTTTCAAAAGCGAGCTC 417
QY 421 tctgaaagcttctcccaaaagcccttggtactaagggtactgtctgtctgctgtg 480
DB 448 TCTGAAAGCTTCTCCCAAAAGACACCTTGGTACTAAGGGTA-----CTGCTG 425
QY 481 ctgctctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 540
DB 496 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 555
QY 541 gtaaaagctccagcccaagaggttcctgcccagaaagccacagccagcagaagcagcgc 600
DB 556 GTAAAAGGCTTCAGCCCGCAGAGGTTCTCTCCCGAAGAACCCAGCGCCAGAAAGCAGCGC 615
QY 601 ctgctccaaagctcagaaggtcctcaaaagctccagcccaagcagaagcactgtctccaaag 660
DB 616 CTGCTCCAAAAGCTCAGAAAGGTCATAAAAGCTCCAGCCCAAGAAAGCACCTGTCTCCAAAG 675
QY 661 catctggcaagaaagacataagtggaatacataaaagtaataaagttcttttaacctg 720
DB 676 CATCTGGCAAGAAAGCATAAAGTGGCAATCAATAAAAGTATAAAGGTTCTTTTIGACCTG 735

QY 721 tt 722
DB 735 tt 737

RESULT 6
LOCUS BC022805
DEFINITION Homo sapiens, clone IMAGE:5426295, mRNA, partial cds.
ACCESSION BC022805
VERSION BC022805.1 GI:18605067
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 713)
Strausberg, R.
Direct Submission
Submitted (04-FEB-2002) National Institutes of Health, Mammalian

```


2/18814/C
LUSIA

ORGANISM

F	COUNT	a	35694	c	35488	q	52460	f	2555	c-hers	
F	COUNT	53326	a	35694	c	35488	q	52460	f	2555	c-hers

10

CONCLUSIONS

2 (pages 1 to 101461) (197101 of 1)

Worley, K. C.

Direct Submission

Submitted (04 - AUG

of Molecular and

Baylor Plaza, Houston, Texas 77030-3099. Tel: 713/798-2100. Fax: 713/798-2101. E-mail: info@baylor.edu

117 1007 '1 idv W0

ac-helphcm.tmc.edu

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84


```
OQ 61 tctcttttgagactcaccggaataatgtgtcgcgaatgtagatgttattgacagagaca 120
|||||
Db 123913 TCTCTTTTGGAGCTTCATGCGGAAATATGTTCGCGATTTAGATGTTATTCATCAGAAACA 123854

OQ 121 gggctttgctgagtgagccttgcaactcaagtgaaggaacagagccatgctttcaagtgca 180
|||||
Db 123853 GGGCTTTGGTCGATGGACCTTGACACTCAAGTGAAGGACAGAGCCATGCTTTCAAGTGCA 123794

OQ 181 tgcagctcaactgatttcatctcctcaagtttctgacagtcgcccacagagatgatctccgac 240
|||||
Db 123793 TGCAGCTCACTGATTTTCATGCTCAAGTTTCGGCAGAGTTCGGCAGAGTTCGGCAGAGTTCGCAAC 123734

OQ 241 aagctgagcaagcagagacacatcaatacaaaatgggcagagcacacgaaggcccaagaaga 300
|||||
Db 123733 AAGCTTGCAAGAGCCAGACATCAATAAATATGGGAGGACACACAGATGGCCCAAGAGA 123674

OQ 301 ttgaagccagagaagaagaagcgaagatgacagatttctgacgttttcaagttatgaag 360
|||||
Db 123673 TTGAAGCCAGAGAAGAAGCAAGCCAGATGACAGATTTTATGCTTTTAAAGTATGAGAG 123614

OQ 361 caaggaataatgaggaacagaataatcaagaatgaagtttaagaagcttcaaaagggcagtc 420
|||||
Db 123613 CAAAGAAACAGGAGACAGATAATCAAGAAATGAAGTTAAGAGCTTCAAAAGCCAGCTC 123554

OQ 421 tctcaaaagcttctcccaaaaagcactgactgactaaggaatgactgctatgctgctgactg 480
|||||
Db 123553 TCTGAAAGCTTCTCCCAAAAAGCCTGTGTACTAAGGGTA-----CTG 123509

OQ 481 ctgctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 540
|||||
Db 123508 CTGCTGNNTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 123449

OQ 541 gtaaaaggctcagagccagagaggttctgcccagagaagccagagccagagccagagccag 600
|||||
Db 123448 GTAAAGGGTTTCCGCCAGAGAGTTCTTCCCGAGAAGCCAGAGCCAGAGCCAGAGCCAGAGC 123389

OQ 601 ctgctccaaagctcagaaggttcaaaagctccagcccaagaagcagctgctcccaag 660
|||||
Db 123388 CTGCTCCAAAGCTCAGAGGGTCAAAAAGCTCCAGCCCAAGAGAGCTGCTGCTGCTGCTGCTG 123329

OQ 661 catctgcaaaaaagcctaagtgagcattcaataaaaaagtaataaagctcttttgacctg 720
|||||
Db 123329 CATCTGCAAGAGCAATAGTGGCAATCAATAAAGTAATAAAGTCTTTTGTGACCTG 123269

OQ 721 tt 722
||
Db 123269 TT 123267

JUL 10
H:U15739
LOCUS HSU15738 701 bp mRNA linear FRI 11-MAY-1999
DEFINITION Homo sapiens CAG-1s1 7 mRNA, complete cds.
ACCESSION U15738 S90520
VERSION U15738.1 GI:508516
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 701)
Aoki,M., Koranyi,L., Riggs,A.C., Wasson,J., Chiu,K.C.,
Vaxillaire,M., Froquel,P., Gough,S., Liu,L., Louis-Keller,H. and
Permutt,M.A.
Identification of trinucleotide repeat-containing genes in human
pancreatic islets
Diabetes 45 (2), 157-164 (1995)
9514156
2 (bases 1 to 701)
Permutt,M.
AUTHORS
DIRECT SUBMISSION
TITLE
Submitted (31-0CT-1994) M. Alan Permutt, Internal Medicine,
Washington University School of Medicine, 660 S. Euclid, St. Louis.
```

MO 63110, USA

COMMENT On May 11, 1998 this sequence version replaced gi:1246101.

FEATURES

Source

1..701

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CAG 7"

/tissue_type="pancreatic islets"

7..648

CDS

/codon_start=1

/product="CAG-1s1 7"

/protein_id="AAC15021.1"

/db_xref="GI:3126984"

/translation="MVRRFVEVGRVAVYFGFHAGKLVAVIDVIDONRALYDGFCTQ
VFOAMPKCMQITDFILKPHSAHQKVROAWKADINTKWAATKRIEAPKKA
KMTDFRKFVKAKKMRNRIKNEVKIKQKALLKASPKKAPGTGSTIAAAAAAAKVF
AKKIATAASKKAPAKVPAOKATGOKNAPAPKQKQKAPAKAPKAPKASGRKA"

repeat=region

452..476

/note="polymorphic region"

/rpt_type=tandem

/rpt_unit=cag

BASE COUNT 225 a 161 c 174 g 140 t

ORIGIN

Query Match

Best Local Similarity 91.1%; Score 657.4; DB 9; Length 701;

Matches 589; Conservative 0; Mismatches 1; Indels 21; Gaps 1;

OQ 12 gcaacacatggttgcaggcgttcgtgagggttgagcgggtgagcgttatgtctctttgga 7;

Db 1 GCAACATGTTGTTTCAAGGCGCTTCGTGGAGGTTGGCGGGTGGCTTATGTCCTCTTIGGA 6;

OQ 72 cctcatcccggaataatggtcgcgattgtagatttattgattgattgattgattgattgattg 131

Db 61 CCTCATGCCGGAATATGTCGCGATTGTAGATGTTATTGATCAGAACAGGCTTTGGTC 12;

OQ 132 gatggaccttgactcaatgagggagagagagagagagagagagagagagagagagagagag 131

Db 121 GATGACCTTTCACCTCAAGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 190

OQ 192 gatttcactcgaatttctgcacagtcgccaccagagatgctccgacagagagagagagag 251

Db 181 GATTTCATCTCAAGTTTCGGCACAGTGCACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240

OQ 252 aaggcagacatcaatacaaaatgggcagccacacacacacacacacacacacacacacacag 311

Db 241 AAGGCAAGACATCAATACAAATGGGCGAGCCACACAGATGGCCAGAGAGATTCAGAGCCAG 300

OQ 312 gaaagaaagcaagatgacagattttgatttttaaaagtattgaagcaagaaagaaatg 371

Db 301 GAAGGAAAGCCAAAGATGACAGATTTTGAATCGTTTAAAGATTATGAAGCAAGAAATG 360

OQ 372 agaacagagataatcaagaatgaagttaagaagcttcaaaagcagcagctctcctgaaagct 431

Db 361 AGGAACAGATAATCAAGAATGAAGTTAAGAGCTTCAAAAGGACAGCTCTCTCTGAAGCT 420

OQ 432 tctcccaaaaagcactgactaaggttactgactgactgactgactgactgactgactgactg 491

Db 421 TCTCCCAAAAAGCAGCTGTGTACTAAGGTA-----CTGCTGCT 450

OQ 492 gctg 551

Db 460 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 510

OQ 552 ccagccagagagttctctcccaagaagccacagcagcagcagcagcagcagcagcagcagc 511

Db 520 CCAGCCCAAGAGTTCTCTGCCAGAAAGCCACAGGCGCAGAAAGCAGCGCTCTCTCCAAA 570

OQ 612 actcagaaggttcaaaaagcctcaaccagcaagcagcagcagcagcagcagcagcagcag 671

Db 580 GCTCAGAGGGTCAAAAAGCTCCAGCCCAAGAGCACCTGCTCTCCAAAGGCTATCTGGCAAG 630

Estimated insert size: 189880; agarose-fp estimation
Estimated insert size: 227960; sum-of-contigs estimation
Quality coverage: 6.18 in Q20 bases; agarose-fp estimation
Quality coverage: 6.82 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 1005: contig of 1005 bp in length
1105: gap of unknown length
1196: contig of 1275 bp in length
2381: gap of unknown length
2481: contig of 1028 bp in length
3509: gap of unknown length
3609: contig of 1511 bp in length
5120: gap of unknown length
5220: contig of 1420 bp in length
5640: gap of unknown length
5740: contig of 1223 bp in length
7963: gap of unknown length
8063: contig of 2507 bp in length
10570: gap of unknown length
10670: contig of 1449 bp in length
12119: gap of unknown length
12219: contig of 2057 bp in length
14276: gap of unknown length
14376: contig of 1824 bp in length
16200: gap of unknown length
16300: contig of 4514 bp in length
20814: gap of unknown length
20914: contig of 5615 bp in length
26528: gap of unknown length
26623: contig of 7418 bp in length
34047: gap of unknown length
34147: contig of 5820 bp in length
39957: gap of unknown length
40056: contig of 5328 bp in length
43394: gap of unknown length
45494: contig of 7856 bp in length
45995: gap of unknown length
53350: contig of 6819 bp in length
53450: gap of unknown length
53551: contig of 6819 bp in length
59664: gap of unknown length
60664: contig of 6353 bp in length
56884: gap of unknown length
56984: contig of 5995 bp in length
73337: gap of unknown length
73437: contig of 8725 bp in length
79333: gap of unknown length
79433: contig of 7154 bp in length
88159: gap of unknown length
88259: contig of 11280 bp in length
95411: gap of unknown length
95511: contig of 9738 bp in length
106792: gap of unknown length
106891: contig of 11303 bp in length
116630: gap of unknown length
116730: contig of 13075 bp in length
128033: gap of unknown length
128133: contig of 19841 bp in length
141208: gap of unknown length
141308: contig of 19841 bp in length
151149: gap of unknown length
151249: contig of 27942 bp in length
189131: gap of unknown length
189290: contig of 4170 bp in length
189291: Location/Qualifiers
1. 230760
/organism="Mus musculus"
/db_xref="taxon:10090"
```

FEATURES
source

```
/clone="RP23-200L5"
/clone_lib="RPC1 mouse BAC library 23"
BASE COUNT 55495 a 57921 c 58057 g 56452 t 2835 others
ORIGIN
Query Match 66.8%; Score 482.2; DB 2: Length 230760;
Best Local Similarity 81.9%; Pred. No. 1.1e-93;
Matches 587; Conservative 0; Mismatches 113; Indels 17; Gaps 2:
Cj 1 cgcctaacgcctcccaacatggttccagcgcttctgaggttgccggtggcgtctatg 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Eb 24765 CGCIGAGCGCGCCCAACATGGTTCAGCGCTTAGTGGAGGTGGCGGGTGCGCTACA 24824
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Eb 24825 TTTCCTTCGGGCCCCCATGCTGGAAGCTGGTCCGAATTGTAGATGTCATTGATCAGAACA 24884
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cj 121 gggctttggtcgatggaccttgacctgaagtagaagaacagccatgcttcaagtca 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24895 GGGCTTTAGTGGATGGACCCCTGTACTCGGTGAGGAGCGCCATGCCCTTCAATGCA 24944
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cj 181 tgcagctcactgatttcatctctcaagtttctgcagtgcccccacagagatgtccgac 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24945 TGCAGCTCAGCTGACTTCAATCTCAAGTTCCACACAGTGCCTCCGCGCAGATATGTCGAA 25004
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cj 241 aagcctggcagaagcagacatcaatacaaatgggagccacacagatgggcccagaaga 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25005 AAGCTTGGGAGAGCGCGATATCAATACGAATGGGAGCGCCACAGATGGGCGCAAGAAA 25064
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cj 301 ttgaagccagagaaaggaagcccaagatgcagatcttgcgtttttaaagtattgaagg 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25065 TTGATGCCAGAGAAAGGAAGCCCAAGATGCACAGCTTTGATCGTTTCAAGGTGATGAAG 25124
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cj 361 caagaaaaatgaggaagcaagaataatcaagaatgaagtttaagaagcttcaaaagcagctc 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25125 CAAAGAAATGAGGAACACAGATAATCAAGACTGAAGTAAAGAAACITTAAGAGAGCTGCTA 25184
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cj 421 tccctgaaagcttctcccaaaaagcaccctgtactaagggtactgctgctgctgctg 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25195 TCCTGAAGGCTTCCTTAAGAAAGCAGCTGTGGCCAA-----AGCTGCCATTG 25232
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cj 481 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 540
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25233 CGGCTGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25292
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cj 541 gtaaaagagctccagccagagaggttctctgccagaagccacagagccagagagagcgc 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25293 GCAAGAGAGCTGCAGAGCCAGAGAGGCTCCTGCTCAGAGGCTGCAGGCGCCAGAGGTGGCAC 25352
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cj 501 ctgctcaaaaagctcaagaaggtcaaaaagctccagccagagagcactgctctccaaagg 560
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25353 CTCTGTCTAAAGGTCAGAAGGCTCAGAAGACCCCGAGAGACCTGCTCTCCCAAGG 25412
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cj 561 catctggcaaaaagcataaagtggaatcaataaaaagtaaaaaggtctctttttgac 717
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25413 CGGCTGGCAGAGAGCGGTGAG-----GCTACACACATAATAAAGTCTTTTTTGAC 25464
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: June 21, 2002, 14:27:35
Job time: 7070 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

*** nucleic - nucleic search, using sw model

Run on: June 21, 2002, 12:36:51 ; Search time 213.27 Seconds
(without alignments)
5812.408 Million cell updates/sec

Title: US-09-923-304-1

Perfect score: 722

Sequence: 1 cgctaagctgcaacatg.....aaagttctttttgacctgtt 722

Scoring table: IDENTITY.NUC

Gafcp 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

*** total number of hits satisfying chosen parameters: 3:72872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1980.DAT.*
- 2: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1981.DAT.*
- 3: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1982.DAT.*
- 4: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1983.DAT.*
- 5: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1984.DAT.*
- 6: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1985.DAT.*
- 7: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1986.DAT.*
- 8: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1987.DAT.*
- 9: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1988.DAT.*
- 10: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1989.DAT.*
- 11: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1990.DAT.*
- 12: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1991.DAT.*
- 13: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1992.DAT.*
- 14: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1993.DAT.*
- 15: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1994.DAT.*
- 16: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1995.DAT.*
- 17: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1996.DAT.*
- 18: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1997.DAT.*
- 19: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1998.DAT.*
- 20: /SIDSI/cgdata/geneseq/geneseq-emb1/HA1999.DAT.*
- 21: /SIDSI/cgdata/geneseq/geneseq-emb1/HA2000.DAT.*
- 22: /SIDSI/cgdata/geneseq/geneseq-emb1/HA2001A.DAT.*
- 23: /SIDSI/cgdata/geneseq/geneseq-emb1/HA2001B.DAT.*
- 24: /SIDSI/cgdata/geneseq/geneseq-emb1/HA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	692.4	95.9	888	AAFI8097	Lung cancer associ
2	567	92.4	709	AAAC00751	Human secreted pro
3	647.6	89.7	820	AAAC00755	Human secreted pro
4	636	88.1	730	AAAC00754	Human secreted pro
5	419.4	58.1	480	AAAC00753	Human secreted pro
6	413.2	57.2	618	AAA16359	Human colon cancer
7	334.8	45.4	409	AAAS7524	cDNA *200 encoding
8	308.8	42.8	503	AAH35325	Human colon cancer
9	295.2	41.0	384	AAH98562	Human ESI-derived

C 10	290.2	40.2	539	22	ABA63242	Human foetal liver
C 11	290.2	40.2	539	22	ABA30476	Probe #8942 for ge
C 12	290.2	40.2	539	22	AAK11710	Human brain expres
C 13	290.2	40.2	539	22	AAK37455	Human bone marrow
C 14	290.2	40.2	539	22	AAI18262	Probe #8195 for ge
C 15	290.2	40.2	539	22	AAI43317	Probe #12003 used
C 16	289.6	40.1	360	21	AAAC00762	Human secreted pro
C 17	272.8	37.8	504	23	AAAS79772	DNA encoding novel
C 18	242.2	33.5	1030	23	AAAS83105	DNA encoding novel
C 19	240.8	33.4	401	22	AAAS33792	Murine translation
C 20	232.4	32.2	1016	23	AAAS79773	DNA encoding novel
C 21	223.8	31.0	395	21	AAAC07628	Human secreted pro
C 22	181.8	25.2	220	22	AAAI5055	Oryctolagus cunicu
C 23	171	23.7	201	16	AAAI19654	Human gene signatu
C 24	145.2	20.2	283	15	AAQ76748	Human genome fragm
C 25	143.5	19.9	335	21	AAAC98326	Human colon cancer
C 26	104	14.4	1662	22	ABAI5326	Human nervous syst
C 27	99.8	13.8	626	21	AAAC50325	Arabidopsis thalia
C 28	99.8	13.8	628	21	AAAC35176	Arabidopsis thalia
C 29	98.2	13.6	867	21	AAAC51997	Arabidopsis thalia
C 30	97.6	13.5	516	21	AAAC47641	Arabidopsis thalia
C 31	97.6	13.5	2232	21	AAAC51844	Arabidopsis thalia
C 32	96.2	13.3	602	23	ABLI10777	Drosophila melanog
C 33	96	13.3	619	21	AAAC40634	Arabidopsis thalia
C 34	95.6	13.2	674	21	AAAC51701	Arabidopsis thalia
C 35	94.6	13.1	160	22	ABAI5690	Zea mays DNA fragm
C 36	94.6	13.1	160	22	ABAI5690	Human foetal liver
C 37	94.6	13.1	160	22	AAK24305	Probe #18750 for g
C 38	94.6	13.1	160	22	AAK50332	Human brain expres
C 39	94.6	13.1	160	22	AAI273394	Human bone marrow
C 40	94.6	13.1	160	22	AAI56302	Probe #17327 for g
C 41	91.2	12.6	756	21	AAAC51892	Probe #24988 used
C 42	80	11.1	485	21	AAAC41024	Arabidopsis thalia
C 43	72.8	10.1	567	21	AAAC29549	Zea mays DNA fragm
C 44	54.2	8.9	499	21	AAAC51643	HIV codon altered
C 45	53.2	8.8	1037	21	AAAS59242	Arabidopsis thalia
					Exons E, C and A c	

ALIGNMENTS

RESULT 1	
AAFI8097	
ID AAFI8097 standard: DNA: 888 BP.	
XX AAFI8097:	
XX AC	
XX 14-MAR-2001 (first entry)	
XX DI	
XX DE	
XX Lung cancer associated polynucleotide sequence SEQ ID 116.	
XX	
KW Human lung cancer associated protein; neuroprotective; cytostatic;	
KW cardioactive; immunomodulatory; muscular active; vulnerar;	
KW gastrointestinal; nephrotropic; antiinfective; gynecological;	
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;	
KW proliferative disorder; wound healing; infectious disease; ds.	
XX Homo sapiens.	
XX WO200055180-A2.	
XX PD	
XX 21-SEP-2000.	
XX PF	
XX 08-MAR-2000: 2000WO-US05918.	
XX FR	
XX 12-MAR-1999: 99US-0124270.	
XX PA	
XX (HUMA-) HUMAN GENOME SCI INC.	
XX PA	
XX (ROSE/) ROSEN C A.	
XX PI	
XX Ruben SM:	
XX WPI: 2000-587514/55.	
DR	

DR. P-PSDB; AAB58221.

PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer.

PS Claim 1; Page 586; 1425pp; English.

Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer associated proteins represented in AAB58105 - AAB58549. Lung cancer associated proteins and polynucleotide sequences, their agonists, and antagonists may have neuroprotective, cytostatic, cardioactive, immunomodulatory, muscular active general, vulnerable, gastrointestinal general, nephrotropic, anti-infective, gynecological, or antibacterial activity. The invention also includes antibodies specific for the protein or polynucleotide sequences. The lung cancer associated polynucleotide sequences may be used for detection of lung cancer. Chromosome identification, as chromosome markers, and for numerous other diagnostic or research purposes. The proteins may be used to treat disorders such as neural, immune, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, and proliferative disorders. The proteins may also be used in the treatment of wounds and infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are used in the course of the invention for the identification and characterisation of the polynucleotide and protein sequences.

Sequence 888 BP; 255 A; 215 C; 236 G; 173 T; 7 other:

```
Query Match          95.9%; Score 692.4; DB 21; Length 888;
Best Local Similarity 99.7%; Pred. No. 5e-155;
```

[illegible]

Q:	617	gaagggtcagaaagactccagcccaagaagacactgcttccaaagagcatctggcgaagaagc	675
Db	735	gaagggtcagaaagactccagcccaagaagacactgcttccaaagagcatctggcgaagaagc	793
Q:	577	ataagtgccaatcatataaaagataaagaagctctctttgacctgtt	722
Db	795	ataagtgccaatcatataaaagataaagaagctctctttgacctgtt	840

RESULT 2

AAC00761

ID AAC00751 standard; cDNA; 709 BP.

AAC00751:

DT 05-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 759.

Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.

CS Homo sapiens.

FN EP1033401-A2.

ED 05-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999: 99US-0122487.

EN (GEST) GENSET.

Dumas Milne Edwards J., Duclert A., Giordano J.

DR WPI: 2000-500381/45.

DR P-PSDB: AAG00755.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures.

The present sequence is one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. An ORF has been identified within the sequence. The 5' ESTs were prepared from total human RNAs or polyA⁺ RNAs derived from 30 different tissues. EST sequences usually correspond reliably to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

Sequence 709 BP; 202 A; 176 C; 183 G; 145 T; 3 other:

Query: Match 92 48. Score 667. PG 21. Length 700.

Best Match	92.43%	Score 697	DB Z1	Length 709
Best Local Similarity	99.63%	Pred. No.	1.5e-159	

Matches	Conservative	Mismatches	Indels	Gaps
Matches 667	Conservative	1	Mismatches 2	Indels 0
Gaps 0				Gaps 0

1 ccctaaccctaccacaatggtatgtccagccacttcattcagagagatttggccgggttggccctatg 50

39 cgcctaacgcctaccacaacatgaatgattccaggccgcttcgttggagggttggccggggtggcctatg 97

51 -c:cccttgcacctcatcccgaaaattggtcgcgaattgtatgattcagaaca 120

DB	98	ttctcttggacctcatgccggaataatggtcgcgattgtagatgttattgattcagaaca	157
OY	121	gggctttggtcgatggaccttgcactcaagtggagagacaggccatgctttcaastgca	180
DF	158	gggctttggtcgatggaccttgcactcaagtggagagacaggccatgctttcaagtga	217
OY	181	tgaagctcacgaatttcatctcaagtttctgcacagtcgccaccagaagaatgctcgac	240
DF	218	tgaagctcacgaatttcatctcaagtttctgcacagtcgccaccagaagaatgctcgac	277
OY	241	aagccttgcaaaaggcagacatcaatacaaaaaggcagccacacagatgggccaaaaga	300
DF	278	aagccttgcaaaaggcagacatcaatacaaaaaggcagccacacagatgggccaaaaga	337
OY	301	ttgaagccagagaaaggaagaagccaagatgacaaatttgaatgcttttaagttataaag	360
DF	338	ttgaagccagagaaaggaagaagccaagatgacaaatttgaatgcttttaagttataaag	397
OY	361	caaaagaaataggaaacagcaataatcaagaataaagcttaagaagcttcaaaaggcagctc	420
DF	398	caaaagaaataggaaacagcaataatcaagaataaagcttaagaagcttcaaaaggcagctc	457
OY	421	tctgaagagcttctccaaaaaaggcactggtactaaggatgactgctgctgctgctg	480
DF	458	tctgaagagcttctccaaaaaaggcactggtactaaggatgactgctgctgctgctg	517
OY	481	ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	540
DF	518	ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	577
OY	541	gtaaaaaggtctccagcccgaaagggttctgtcccgaaagccacaggtcagaagcagcgc	600
DF	578	gtaaaaaggtctccagcccgaaagggttctgtcccgaaagccacaggtcagaagcagcgc	637
OY	601	ctgtctcaaaagcttcagaaggttcaaaaagcttcagcccgaaagacacctgtgtccaaag	660
DF	638	ctgtctcaaaagcttcagaaggttcaaaaagcttcagcccgaaagacacctgtgtccaaag	697
OY	661	catctggcaa	670
DF	698	catctggcaa	707

RESULT 3
 AA:00765
 IE AAC00765 standard: cDNA: 820 BP.

AAC00765:

DT 05-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 763.

KW	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW	gene therapy; chromosome mapping; ss.

QC Homo sapiens.
XX EP1033401-A2.
PN

05-SEP-2000.

XX: 21-FEB-2000: 2000EP-0200510.

XX
PF 26-FEB-1992: 99JIS-0122487

PA (GEST) GENSET.

Pl Dumas Milne Edwards J, Duclert A. Giordano T.

AA.
DE
WPI; 2000-500391/45..

DF P-PSDB: FAG00759.

XX PI PI PI XX PS XX CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC XX SQ

Query Match	89.73;	Score 647.6;	DB 21;	Length 820;
Best Local Similarity	98.93;	Pred. No. 1.4e-154;		
Matches 647:	Conservative	3;	Mismatches	4;
			Indels	0;

03 17 catggatgttcaggacacttcattggaagattggccgaatggcctatgtctcctttggacctca -5
|||||
155 catgaatattcaggacacttcattggaagattggccgaatggcctatgtctcctttggacctca -35

Q: 77 +cccggaatat+ggtccgcatctgtataatgttatttgacagaaacagggctttggtcgatgg
|||||
D: 225 +cccggaatat+ggtccgcatctgtataatgttatttgacagaaacagggctttggtcgatgg
|||||

Q: 137 accttgactcaatgaggagacaggccatgaccttcaagtgcattgcagctcacctgattt 139
|||||
A: 137 accttgactcaatgaggagacaggccatgaccttcaagtgcattgcagctcacctgattt 139
|||||
Q: 185 accttgactcaatgaggagacaggccatgaccttcaagtgcattgcagctcacctgattt 187
|||||
A: 185 accttgactcaatgaggagacaggccatgaccttcaagtgcattgcagctcacctgattt 187
|||||

Q: 197 catcctcaagt tttgtcacagtgccaccacaaatattgtccgaagaacctgacaaggc 256
||||| ||||| ||||| ||||| : ||||| ||||| ||||| ||||| ||||| ||||| |||||

257 agacatcaatacaaaatgggcagccacacacgtgggccaagaattggaagccagaaag 316

317 gaaagcagaatgacagattttgatcgttttataaagtattgaaggcaagaaaaatgaggaa 376

CC 495 GAAAGGCAAGATGmKGAATTTTGTATCGTTTAAAAGTATTGAAGGCAGAAGAAAAAGAGaa 527

Q: 377 CAGATAATCAGAAATGAAGTTAAGAGCTTCAAAAAGCAGCTCTCTCTCTGAAAGGCTTCTC 436

DG 525 CAGGAATACAAAGCATGAAGTTAAGAAGCTTCAAAAAGGCAGCTCCTCCGTGAAGCTTCGG
Q: 437 CAATAAACCGCACCTGGTACTTAAGGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

D6 585 caaaaagcacctggtaactaaaggatcgtgctgctgctgctgctgctgctgctgc 614

D7 497 tgcctgctgctaagtccagcaaaaaagatcacgccgcgagtaaaaaagctccacg 555

DB 645 TGCCTGCTGCTAAAGTCCAGCAAAAAGATACCGCCGCGAGTAAAAAGGCTCCAGC 704

Q: 557 CCAAGAGGCTCCTGCTCCAGGAAAGGCCACAGGCGAGGAGGCTGCTCCAAAAAGCTCA 616

705 ccagaaagtctctgccagaaagcacaaagccagaaagcagcctgctcnaaaagctca 754
617 gaaggtccaaaaagctccaaccagaaagacacctctccaaaagcatctctgcaa 570

DB 55 gaaggtcagaaagctccagcccaagaagcacctgtccaaaggatctggcaa 818

AA557524
ID AAS57524 standard; cDNA; 409 BP.
AC AAS57524;
XX 13-FEB-2002 (first entry)
XX cDNA #200 encoding portion of a human colon tumour protein.
XX Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.
XX Homo sapiens.
XX WO200173027-A2.
XX 04-OCT-2001.
XX 22-MAR-2001; 2001WO-US09246.
XX 24-MAR-2000; 2000US-191597P.
XX 04-MAY-2000; 2000US-202024P.
XX 05-MAY-2000; 2000US-202189P.
XX (CORI-) CORIXA CORP.
XX Meagher MJ, Xu J, King GE.
XX WPI: 2001-611627/70.
XX New colon tumour proteins and related nucleic acid, useful for
XX treatment, prevention, diagnosis and monitoring of cancer.
XX Claim 4: Page 91: 299pp; English.
XX Th present invention relates to the isolation of novel cDNA sequences
XX encoding for at least an immunogenic portion of human colon tumour
XX proteins. The sequences of the invention are useful in pharmaceutical
XX compositions and vaccines for the prevention and treatment of cancers
XX such as colon cancer. They are also useful for the diagnosis and
XX monitoring of such cancers. Antibodies to the colon tumour proteins
XX and antigen presenting cells that express polynucleotides encoding
XX colon tumour proteins can be used to inhibit the development of
XX cancers. T-cells that react specifically with colon tumour proteins
XX are useful for removing tumour cells from samples (e.g. blood) and
XX for cancer treatment. The polynucleotides sequences are also useful in
XX gene therapy. AAS57525-AAS5880 represent the cDNA sequences of the
XX invention that encode for portions of human colon tumour proteins.
XX Sequence 409 BP: 100 A; 100 C; 116 G; 91 T; 2 other;

Query Match 46.48; Score 334.8; DB 23; Length 409;
Best Local Similarity 98.8%; Pred. No. 2.4e-75;
Matches 336; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 8 cgcgtcccaacatggttccagcgtcttcgtggaagttggcggagtgccctatctctt 57
Dh 70 cgcgcgaacatggttccagcgtcttcgtggaagttggcggagtgccctatctctt 129
Qy 68 tggaccttcacgcggaaattggtgcgagttgtagatgttatgacagacaggcttt 127
Dh 130 tggaccttcacgcggaaattggtgcgagttgtagatgttatgacagacaggcttt 189
Qy 128 ggtcgatggaccttgcactcaagtgaagacagagggccatcctttcaagtgcagct 197
Dh 190 ggtcgatggaccttgcactcaagtgaagacagagggccatcctttcaagtgcagct 249
Qy 188 cactgatttcctcactcaagttcttcacagtgccaccacaaagtatctccgacaagctg 247
Dh 250 cactgatttcctcactcaagttcttcacagtgccaccacaaagtatctccgacaagctg 309
Qy 248 gcaagaagcagacatcatcaataaataaggcagccacacacatgggcccagaagattgaagc 307

Db 310 gcagaaggcagacatcncatacaaaatggcagccacacgatggcccaagaagattgaagc 369
Qy 308 cagagaagaagaacccaagatgacagattttgatcgtttt 347
Dh 370 cagagaagaagaacccaagatgacagattttgatcgtttt 409
RESULT 8
AAH35325
ID AAH35325=standard; cDNA; 503 BP.
AC AAH35325;
XX 03-SEP-2001 (first entry)
XX Human colon cancer antigen encoding cDNA SEQ ID NO:2407.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; ss.
XX Homo sapiens.
XX WO200122920-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US26524.
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI: 2001-235357/24.
XX P-PSDB: AAG75920.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 1: Page 3958; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated P,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAG77789 represent sequences used in the exemplification of the
XX present invention.
XX N.B. Pages 665 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX Sequence 503 BP: 182 A; 120 C; 96 G; 96 T; 9 other;

Query Match 42.8%; Score 308.8; DB 22; Length 503;
Best Local Similarity 94.1%; Pred. No. 1e-68;
Matches 354; Conservative 0; Mismatches 11; Indels 11; Gaps 3;
Qy 254 ggcagacatcaatacaaaaatggcagccacacgatggcccaagaagattgaagcagaga 313
Dh 13 ggcagacatcaatacaaaaatggcagccacacgatggcccaagaagattgaagcagaga 72

OY 314 aaggaagcagaatgacagatttgcctttttaaagttatgaagcgaagaaatgag 373
 Db 73 aaggaagcagaatgacagatttgcctttttaaagttatgaagcgaagaaatgag 132
 OY 374 gaacagataatacaagaataaagttcaagaagcttcaaaagcagctctctgaagcttc 433
 Db 133 gaacagataatacaagaataaagttcaagaagcttcaaaagcagctctctgaagcttc 192
 OY 434 tcccaaaaaagcaccctgggtactgaaggtactgctgctgctgctgctgctgctgctg 493
 Db 193 tcccaaaaaagcaccctgggtactgaaggtactgctgctgctgctgctgctgctgctg 243
 OY 494 tgcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 553
 Db 244 tgcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 303
 OY 554 agcccaagaagttcctgcccagaaagccacagagccagaaagccagcctctccaaagc 613
 Db 304 cacccaagaagttcctgcccagaaagccacagagccagaaagccagcctctccaaagc 361
 OY 514 tcagaaggggtcaaaaa 629
 Db 362 tcanaaaggtcaaaaa 377
 RESULT 9
 AAH98682 ID AAH98682 standard; cDNA: 384 BP.
 AC AAH98682:
 DI 12-OCT-2001 (first entry)
 XX Human EST-derived coding sequence SEQ ID NO: 539.
 DE Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 OS Homo sapiens.
 XX WO200154477-A2.
 PN 02-AUG-2001.
 XX 25-JAN-2001: 2001WO-US02687.
 PR 25-JAN-2000: 2000US-0491404.
 PR 17-JUL-2000: 2000US-0617746.
 PR 03-AUG-2000: 2000US-0631451.
 PR 15-SEP-2000: 2000US-0663870.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen F, Asundi V;
 PI Cao Y, Dirmacac PA, Zhang J, Werhman T;
 XX WPI: 2001-476164/51.
 DP P-FSDB: AAM24023.
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PI antibodies and research use
 XX Claim 1: Page 550: 1275pp; English.
 PS The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess

CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.
 XX Sequence 384 BP: 124 A; 80 C; 84 G; 96 T; 0 other;
 SO
 Query Match 41.0%; Score 296.2; DB 22; Length 384;
 Best Local Similarity 86.1%; Pred. No. 1.4e-65;
 Matches 328; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
 OY 30 cgtctctgaaggttgccgggtggcctatgtctcttcttgacctcatgcggaaaattg 89
 Db 3 cgtctctgaaggttgccgggtggcctatgtctcttcttgacctcatgcggaaaattg 62
 OY 90 gtgcgaattgaagatttattgatcaacaagggcttggctgcatgagaccttgcaactca 149
 Db 53 ctgagattgaatgtccagatcatgcaacaccttggctcaatgagaccttgcaactca 122
 OY 150 gtgaggagacagggcctcttcaagtgcagtcagctcactgatttctcctcaagt 209
 Db 123 gtgaggagacagggcctcttcaagtgcagtcagctcactgatttctcctcaagt 182
 OY 210 ctgcacagtcgccacacagaggtatgtccgaagcctggcagagggcagacatcata 269
 Db 183 tcgcacagtcgccacacacagaggtatgtccgaagcctggcagagggcagacatcata 242
 OY 270 aaatgggcaagccacagatggcccaagaagattgaagccagaaagaaagcaaatg 329
 Db 243 atatgggcaagccacacagatggcccaagaagattgaagccagaaagaaagcaaatg 302
 OY 330 acgattttgcctgttttaagttatgaagcgaagaaataatgaggaacataatcaag 389
 Db 303 acgattttgcctgttttaagttatgaagcgaagaaataatgaggaacataatcaag 352
 OY 390 aatgaacttaagaagcttcaa 410
 Db 363 aatgaacttaagaagcttcaa 383
 RESULT 10
 ABA63242/C
 ID ABA63242 standard; DNA: 539 BP.
 AC ABA63242:
 XX 01-FEB-2002 (first entry)
 DI Human foetal liver single exon nucleic acid probe #11547.
 DE Human: foetal liver; gene expression; single exon nucleic acid probe; ss.
 OS Homo sapiens.
 XX WO200157277-A2.
 PN 09-AUG-2001.
 XX 30-JAN-2001: 2001WO-US00669.
 PR 04-FEB-2000: 2000US-0180312.
 PR 26-MAY-2000: 2000US-0207456.
 PR 30-JUN-2000: 2000US-0608408.
 PR 03-AUG-2000: 2000US-0632366.
 PR 21-SEP-2000: 2000US-0234687.
 PR 27-SEP-2000: 2000US-0236359.
 PR 04-OCT-2000: 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2001-483447/52.
 XX


```

OY 188 cactgatttcctcaagatttctgcacagtgccca--ccagaagatatccgacaagcc 245
DB 237 CACTGACTTTATCTTCAAGTTCACCCACCCAGTCCGCCACCCAGAGCCATGGCCCAAGAGGCC 178
OY 246 tggcagaagagagacatcaatacaaaaatgggcagccacacacagatggcccaaaagattgaa 305
DB 177 TGGCAGAAGGCGCAACCATCAATACAAATGGACAGGCACACAATGGCCCAAGAGAGTAA 118
OY 306 gccgagagaaggaagcccaagatcacagatttctgcagtttcaagtttaagttatgaagcacaag 365
DB 117 GCCAAAGAGGGGAGCCCAATATAGATATTGATCAATTTT--AGTCATGAAGGCAAG 50
OY 365 aaaaatgaggaacagataatacaagaatgaagtttaagagcttcaaaagcagctctcct 424
DB 59 AAAATGAGGAACAGATGATCAAGATTAATAATGAAGAGCTTCAAAAGCTGCTGCTCT 1

RESULT 14
AA118262/C
ID AA118262 standard: DNA: 539 BP.
AC AA118262:
CT 12-OCT-2001 (first entry)
DE Probe #8195 for gene expression analysis in human cervical cell sample.
KW Probe: human: microarray: gene expression: cervical epithelial cell;
XX cervical cancer; ss.
XX Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001: 2001WO-US005070.
XX 04-FEB-2000: 2000US-0180312.
XX 26-MAY-2000: 2000US-0207456.
XX 30-JUN-2000: 2000US-0608408.
XX 03-AUG-2000: 2000US-0632356.
XX 21-SEP-2000: 2000US-0234687.
XX 27-SEP-2000: 2000US-0236359.
XX 04-OCT-2000: 2000US-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR:
WPI: 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for
PI analyzing gene expression in human cervical epithelial cells -
XX
FS Claim 25: SEQ ID No 8195: 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIFO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 539 BP: 130 A; 142 C; 116 G; 151 T; 0 other;

```

Query Match 40.2%; Score 290.2; DB 22; Length 539;
 Best Local Similarity 85.7%; Pred. No. 5.4e-64;

```

Matches 359; Conservative 0; Mismatches 53; Indels 7; Gaps 3;
OY 8 cactgccaacatgctgttccaggcgcttcgtgaggttgccggtgacctatgtctcctt 57
DB 414 CGCCACCACCAAGGTGTTTCAGGCACCTTTGTGGAGTTGGCAAGTGGCTTACATCTC--- 356
OY 68 tggacctcatgcccgaataattgctgcgattgtagatgttattgaatcagacaagggcttt 127
DB 257 TGGACCTTCATGCTGGNAAGCTGGTTGGCATTTAGATTTGTTTGTATGAGAACAGGGCTTT 298
OY 128 ggtcgtatggaccttcacactcaagtgaaggagacagggccatgccttttcaagtgcagct 187
DB 297 GGTGTGTGAGCTTGCCTCAGTCAAGTAAAGGACAGAGCCATGCCATTTAAATGATGAGCT 238
OY 188 cactgatttctcctcaagtttctgcacagtgcga--ccagaagtatgtccacaagcc 245
DB 237 CACTGACTTTATCTTCAAGTTCACCCAGTCCGCCACCCAGAGCCATGGCCCAAGAGGCC 178
OY 246 tggcagaagagagacatcaatacaaaaatgggcagccacacacagatggcccaaaagattgaa 305
DB 177 TGGCAGAAGGCGCAACCATCAATACAAATGGACAGGCACACAATGGCCCAAGAGAGTAA 118
OY 306 gccgagagaaggaagcccaagatcacagatttctgcagtttcaagttttaaagttatgaagcacaag 365
DB 117 GCCAAAGAGGGGAGCCCAATATAGATATTGATCAATTTT--AGTCATGAAGGCAAG 50
OY 365 aaaaatgaggaacagataatacaagaatgaagtttaagagcttcaaaagcagctctcct 424
DB 59 AAAATGAGGAACAGATGATCAAGATTAATAATGAAGAGCTTCAAAAGCTGCTGCTCT 1

RESULT 15
AA143317/C
ID AA143317 standard: DNA: 539 BP.
XX
AC AA143317:
CT 17-OCT-2001 (first entry)
DE Probe #12003 used to measure gene expression in human placenta sample.
KW Probe: microarray: human: placenta: prenatal diagnosis;
XX genetic disorder; ss.
XX Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001: 2001WO-US006663.
XX 04-FEB-2000: 2000US-0180312.
XX 26-MAY-2000: 2000US-0207456.
XX 30-JUN-2000: 2000US-0608408.
XX 03-AUG-2000: 2000US-0632356.
XX 21-SEP-2000: 2000US-0234687.
XX 27-SEP-2000: 2000US-0236359.
XX 04-OCT-2000: 2000US-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR:
WPI: 2001-48897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PI analyzing gene expression in human placenta -
XX
FS Claim 25: SEQ ID No 12003: 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for

```

CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.

S0 Sequence 539 BP: 130 A; 142 C; 116 G; 151 T; 0 other:

```
Query Match      40.2%; Score 290.2; DB 22; Length 539;
Best Local Similarity 85.7%; Pred. No. 5.4e-64;
Matches 359; Conservative 0; Mismatches 53; Indels 7; Gaps 3:

Qy 8 cgcgcacacatagtgttcaggcgcttcgtgagggttgccggggtgagccctatctctctt 67
Db 414 CGCACCAACAATGIGITCAGGCACITITGGAGGTTGGCCAAAGTGGCTTACATCTC--- 356

Qy 58 tggacctcatgcccggaataatggtcgcgattgttagatgttatgatcagaagagggcctt 127
Db 357 TGGACCTTCATGCTGGAAAGCTGGTTCGGATTATAGATGTTGTTGATCAGAGAAAGGCGCTT 298

128 gctcgatggaccttgacctcaaatgaggagagacagggccatgctcttcaagtgttgaagct 187
Db 297 GGTGGTGGAGCTTGCACTCAGGTAGGAGACAGCGCCATGCCATTAAATGCCATGCCAGCT 238

Qy 188 cactgatctcctcaagtctctgcacagtccca---ccagaagtatgctcacaagcc 245
Db 237 CACTGACTTTTATCTCTCAAGTTCCTCACCCAGTGGCCACCCGAGGCGATGGCTAAGAGGCC 178

Qy 245 tggcagaagcgagacatcaatacaaaaatgggcagccacacgattgggcccaagaaattgaa 305
Db 177 TGGCAGAAGGCAACCATCAATACAAAATGGACAGGCGACACAAATGGGCGCAGAAAGIGCAA 118

Qy 305 gccagagaaggaaagccaaagatgcagagatttgaatcgttttaaaagttatcagaagcaag 365
Db 117 GCCAAGAGAGGGGAAGCAAAATCATAGATTTTGATCATTTT--AGTCATCAATGACAAAG 60

Qy 365 aaaaagaagaaataatcaagaatgaagttaaaagcctcaaaaagcgagctctct 424
Db 59 AAAATGAGGAACAGAAATGATCAAGAATTAAATTAAGGAAGCTTCAAAAGGCTGTGTGCTCT 1
```

Search completed: June 21, 2002, 14:24:21
Job time: 5450 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Pun on: June 21, 2002, 11:55:00 : Search time 1700.22 seconds
(without alignments)
5731.157 Million cell updates/sec

Title: US-09-923-304-1
Perfect score: 722
Sequence: 1 cgctaagctgcaacatg.....aaggttttttttgcctgttt 722

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_estl:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed. and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	698.2	96.7	755	10	BG756624
2	693.2	96.0	822	10	BG763626
3	691.8	95.8	730	10	BG109146
4	690.2	95.6	901	10	BG748004
5	688.4	95.3	797	10	BM083723
6	688.0	94.3	728	9	AU123120
7	680.4	94.2	818	9	AL531631
8	679.8	94.2	723	10	BM472466
9	679.9	94.0	730	10	BG396368
10	677.4	93.8	746	10	B1914072
11	675.2	93.5	732	10	BG260421
12	674.4	93.4	705	10	BM043710
13	673.4	93.3	712	9	AL528184
14	671.4	93.0	1116	10	BM480399
15	670.9	92.8	724	10	BG437749
16	668.4	92.6	711	10	BM478504
17	668.2	92.5	732	10	BM009159

18	555.6	92.3	673	10	BG774138
19	556.2	92.3	752	10	BM423398
20	554	92.0	723	10	BM019074
21	554	92.0	845	10	BG473548
22	553	91.8	769	10	B1856897
23	551.8	91.7	716	10	BG764388
24	551.2	91.6	728	10	BG485150
25	550.8	91.5	749	9	AL583702
26	550.6	91.5	807	9	AL518218
27	550.4	91.5	708	9	AL526194
28	552.8	91.4	778	10	B1252458
29	559	91.3	748	10	BG282209
30	558.8	91.2	715	10	BM471385
31	558	91.1	876	10	BG489161
32	556.2	90.9	808	10	BG751023
33	555.4	90.8	1166	10	BM466808
34	554.4	90.6	598	10	B1092706
35	553.6	90.5	840	10	BF663213
36	553.4	90.5	725	9	AL527211
37	553.4	90.5	1122	10	BF663343
38	553.2	90.5	719	10	BG759156
39	553.2	90.5	760	10	B1114711
40	552.6	90.4	740	10	BG503437
41	551.8	90.3	988	10	B1520381
42	551.6	90.2	753	10	BF965976
43	551.4	90.2	1206	10	BM458518
44	551.2	90.2	733	10	BG496235
45	550.6	90.1	837	10	BF036445

ALIGNMENTS

RESULT 1

BG756624 755 bp mRNA linear EST 15-MAY-2001
LOCUS 602713776F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4853825 5'
DEFINITION mRNA sequence.
ACCESSION BG756624
VERSION 1
KEYWORDS ESI.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 755)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs.r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LCM1700 row: h column: 19
High quality sequence stop: 718.
Location/Qualifiers
1. .755

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4853826"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library.

BASE COUNT 247 a 168 c 191 g 149 t

OPIGIN

Query Match 96.7%; Score 698.2; DB 10; Length 755;
Best Local Similarity 99.6%; Pred. No. 7e-138;
Matches 700; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Oy 12 gcaacacatggtttcagcagcttcgtgaggttgccgggtggcctatgtctctcttggga 71
Db 2 GCCAACATCGTGTTCAGCGCGTTTCGAGGTTGGCGGGTGGCTATGCTCTCTTGGGA 51

Oy 72 cctcatgccgaaattgctgcacatgtatcatatttattgacacagagggcttgcctc 131
Db 62 CCTCATGCCGNAATTCGTCGATGTAGATGTATTGATCATACACAGCGCTTTCGTC 121

132 gatgaaccttgactcaagtgagagacagccatgccctttcaagtgcacacagctcact 191
122 GATGACCTTCACATCAAGTCAGGAGACAGGCCATGCTTTCATAGTCATGACGTCAC 181

Oy 192 gatttcacctcaagttctcacagtcgccacagagatgtatccacaaagcctgcag 251
Db 182 GATTTCATCTCAAGTTTCCGACAGTGCACCACAGAGATGTGTCGACAAAGCTGCGAG 241

Oy 252 aaggcagacatcaatcacaaatgggcagccacacagatggcccaagaagattgaagccaga 311
Db 242 AAGGCAGACATCAATACAAATGGCGAGCCACACAGATGGGCCAASAGATTAAGGCCAGA 301

Oy 312 gaaggaaagccaagatcaacagatttgatgttttaadttatgaaggcaagaaatg 371
Db 302 GAAGGAAAGCCAGATGACAGATTTTGATCGTTTAAAGTTATGAAGCCAAAGAAATG 361

Oy 372 agaacagaataatcaagaatgaagtttaacagcttcaaaagacacctctctcgaagct 431
Db 362 AGGAACAGAAATATCAGAAATGAAGTTAAGAGCTTCAAAAGGCGAGCTCTCTGAAAGCI 421

Oy 432 tctcccaaaaagcactgtactaaaggttactgtctgtctgtctgtctgtctgtctgtct 491
Db 422 TCTCCCAAAAAGCCTGTGTACTAAGGTGACTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 481

Oy 492 gctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtct 551
Db 482 GCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 541

552 ccagcccaagaaggttctgtcccaagaagccacagggccagaaagcagcgcgtgtctccaaa 611
542 CCAGCCCAAGAAGGTCTGTCCCAAGAAGCCACAGGGCCAGAAAGCAAGCGCTGTCTCCAAA 601

Oy 612 gctcagaagggtcataaagctccagcccaagaagaccctgtctcgaagcactctgaag 671
Db 602 GCTCAGAAGGGTCANAAAGCTCCAGCCCAAGAAGACCTGTCTCCAAAGGTCATCTGCAAG 661

Oy 672 aaagcataagtggaactcataaaagaataataaagggttctctttt 714
Db 662 AAAGCATAAGTGGCAATCATAGAAGTAATAGAGGTTCITTTTT 704
```

RESULT 2

LOCUS

DEFINITION BG763626 822 bp mRNA linear EST 15-MAY-2001
602735840f1 NIH_MGC_45 Homo sapiens cDNA clone IMAGE:4861305 5'

mRNA sequence.

ACCESSION

VERSION BG763626

KEYWORDS BG763626.1 GI:14074279

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHOR
TITLE
JOURNAL
COMMENT

1 (bases 1 to 822)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: ccapbs@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1719 row: P column: 10
High quality sequence stop: 794.

FEATURES

source

1..822
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4861305"
/clone_lib="NIH_MGC_45"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTR7; Site:1; XhoI: Site:2;
EcoRI: cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 250 a 179 c 206 g 177 t

ORIGIN

Query Match 95.0%; Score 693.2; DB 10; Length 822;
Best Local Similarity 99.2%; Pred. No. 8.1e-137;
Matches 708; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

```
Oy 9 gctgccacatggtttcagcagcttcgtgaggttgccgggtggcctatgtctctctt 58
Db 2 GCCGCAACATGGTGTTCAGGCGCTTCGTGGAGGTGGCGGGTGGCTATGCTCTCTTT 61

Oy 69 ggaacctcatgccggaaaaattgctgcgattgtagatgttattgatcacacagagcttg 128
Db 62 GGACCTCATGCCGAAAAATTGTCGCGATTGTAGATGTTATTGATCAGACAGGCTTIG 121

Oy 129 atcgatgaaccttgacactcaadtgaaggagacagggccatgcctttcaagtcacagctc 188
Db 122 GTCGATGAGCCCTTGCACITCAAGTAGGAGACAGGCCATGCGCTTTTCAGTGCATGCAGCTC 181

Oy 189 actgatttcactcctcaagtttctgcacagtgccaccagaaagtatgctgcgaagcctgg 248
Db 182 ACTGATTTTCCTCAAGTTTCCGCGACAGTGCCACCACAGAGATGTCGCCACAGGCTGG 241

Oy 249 cagaaggcagacatcaatacaaaaatggcgagccacacacatgggccaagaatgaagcc 308
Db 242 CAGAAGGCAGACATCAATACAAAATGGCGCAGCCACACAGATGGCGCAGAAAGATTGAAGCC 301

Oy 309 agagaaagaaagcccaadatatcaadattttgatcgtttttaaagttatgaaggcaagaaa 358
Db 302 AGAGAAAGGAAGCAACAGATGACAGATTTTCATCGTTTAAAGTTATGAGGCCAAGAAA 351

Oy 359 atgaggaacaataataatcaagaatgaagtttaagaagcttcaaaaggcagctctcctgaaa 428
Db 362 ATGAGGAACAGAAATTAATCAAGAAATGAAGTTAAGAAGCTTCAAAAGGCGAGCTCTCTGAAA 421

Oy 429 gcttctcccaaaaagccactctgactaagaggtactgctgctgctgctgctgctgctgct 488
Db 422 GCTTCTCCCAAAAAGACACCTGGTACTNANGGTA---CTGCTGCTGCTGCTGCTGCTGCT 478

Oy 489 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 548
```



```
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library.
1-

BASE COUNT      288 a      217 c      240 g      156 t
ORIGIN

Query Match      95.5%; Score 690.2; DB 10; Length 901;
Best Local Similarity 98.7%; Pred. No. 3.5e-136;
Matches 706; Conservative 0; Mismatches 8; Indels 1; Gaps 1:

Q: 9 gctgccacatggtgttcaggcgttcgtgaggttgccggatgacctatgtctcttt 68
|||||
2 GCTGCCACATGTTTCAGCGCTTCGTGGAGTTGGCCGGTGGCCATGTCCTTT 61
|||||
59 ggaacctatccgggaaattggtcgagttgagatgtattatcagaaacagggttg 129
|||||
62 GGACCTCATGCCGAAATTCGTCGAGTTGTAGTGTATTGATCAGAACAGGGCTTG 121
|||||
129 gtcgatgaccttcactcaagtgaagaaagacagccatgcctttcaagtcatcagctc 189
|||||
122 GTCATGGACCTTTCACCTCAAGTGAGGAGACAGGCCAAGCTTTCAGTGCATGCAGCTC 181
|||||
189 actgatttcacctcaagttcttcagcagtgccaccagagatgtgccacaagcttg 248
|||||
182 ACTGATTTTCCTCAAGTTTCGCCACAGTGGCCACCAGAAAGTATGTCCACAAAGCTTG 241
|||||
249 cagaagcagacatcaatacaaaatggcgagccacacagatggccaaagaagtgaagcc 308
|||||
242 CAGAAGCGACACATCAATACAAATGGCAGCCACACACGATGGCCCAAGAAATGTAAGCC 301
|||||
309 agaaaaggaaagcagaatgacagattttgatctgttttaaaattatgaaggcaaaagaa 358
|||||
302 AGAAAGGAAAGCCCAAGATGACAGATTTTGATCGTTTAAAGTTATGAAGGCCAAGAAA 361
|||||
359 atgagaacagaataatcaagaatgaagttaagaagcttcaaaaggcagctctctctgaaa 428
|||||
362 ATGAGGAACAGATAATCAAGAAATGAAGTTAAGAAGCTTCAAAAGGCAGCTCTCTTGAAA 421
|||||
429 gcttctcccaaaaagcaccctgactactaaggtactactactactactactactact 488
|||||
422 GCTTCTCCCAAAAAGACACTGTGTTACTAAGGGTACTGCTGCTGCTGCTGCTGCTGCT 481
|||||
489 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 548
|||||
482 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 541
|||||
549 gctccagccagaaaggttctgcccagaaagccacagagccagaaagcagc-gcctgctcc 607
|||||
542 GCTCCAGCCAGAAAGGTTCTGCTCCAGAAAGCCACAGGCGCAGAAAGCAGCGGCTGCTCC 601
|||||
608 aaaagctcagaagggtcaaaaagctccagcccgagaaagcaccctgcttcaaaagcattctg 667
|||||
602 AGAAGCTCAGAAAGGTCAAAAGCTCCAGCCCAAGAAAGCACTGCTCCAAAGGCATCTGG 661
|||||
658 cagaaagcagaatgagcaatcaataaaagtaataaaggttcttttgacctgtt 722
|||||
662 CAAGAAAGCATAGTGGCAATCATACGCAGTAATAGAGGTTCTTAATGAGCTGTT 716
|||||

RESULT      5
LOCUS      BM083723
DEFINITION      image92c_2-2001/smk340bdf41.x1 NIH_MGC_16 Homo sapiens cDNA clone
IMAGE:4548493 5'. mRNA sequence.
ACCESSION      BM083723
```

```
BM083723.1 GI:16951354
ESI.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 797)
Kale,P.I., Harsch,I.J., Polta,P.A., Nelson,D.O., Sanders,C.G. and
Prange,C.K.
The I.M.A.G.E. Consortium quality control effort: clone
resequencing for verification
Unpublished (2001)
Other_ESIs: BG474329
Contact: Prange CK
The I.M.A.G.E. Consortium
Lawrence Livermore National Laboratory
Livermore, CA, USA
Email: help@image.llnl.gov
This read has been verified (found to hit its original self in the
correct orientation), as part of the I.M.A.G.E. Consortium quality
control effort. High quality sequence is defined as having 100 or
more base pairs with a phred quality value of 20 or greater, where
a sliding window of 4 base pairs with a phred quality value of 15
or greater marks the beginning and end of the sequence. For
information on obtaining this clone, please contact
info@image.llnl.gov.
Plate: LLCM1427 row: e column: 4
Seq primer: -21ml3
High quality sequence stop: 797.
Location/Qualifiers
1..797
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORB7; Site: 1; XhoI; Site: 2;
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
6 others

BASE COUNT      241 a      174 c      200 g      176 t
ORIGIN

Query Match      95.3%; Score 688.4; DB 10; Length 797;
Best Local Similarity 99.4%; Pred. No. 8.4e-136;
Matches 700; Conservative 0; Mismatches 3; Indels 1; Gaps 1:

Q: 20 gctgttcaggcgttcgtgagattgagccaggtgagcctatgtctcttggacctatgc 79
|||||
Db 1 GGTGTTTCAGCGCTTCGTGGAGGTGGCCGGTGGGCTATGCTCTCTTTGGACCTCATGC 60
|||||
80 cgaataattggtcgctgattatagatttatgatcagaacagggtttgtgcataggacc 130
|||||
Db 61 CGGAAATTTGGTCGCGATTGTAGATGTTATGATCAGAACAGGSCITTTGGTCGATGACC 120
|||||
Q: 140 ttgactcaagttgagagacagagccatgcctttcaagtgcatacagctactaatcat 190
|||||
Db 121 TTGCAGTCAAGTGAGGAGACAGGCCATGCTTTCAGTGTGATGCAGCTCACTGATTTCAT 180
|||||
Q: 200 cctcaagtttctacagtgccaccagaaagtatgccacaagcctgcagaaagcaga 259
|||||
Db 181 CTTCAAGTTTCCACAGATGCCCCCAGAAAGTATGTCGACAAAGCCTGGCAGANGCAGA 240
|||||
Q: 260 catgaatacaaaatggcgagccacacagatggccagaagaattgaagccagaaagaaa 310
|||||
Db 241 CATCAATACAAATGGCAGCCACACAGATGGGCCAAGAGATTGAAGCCACAGAAAGGAA 300
|||||
```

```
QY 320 agccaaatcacagatttttgcgttttaaaagtattgaaggcaaaaataatgaggaaacag 379
|||||
Db 301 AGCCAAATGACAGATTTTGAATGTTTAAAGTTATCAAGGCAAGAAATAGGGAACAG 360
QY 380 aataatcaaaatgaatttaagaagcttcaaaaggcagctctctctgaagaagctctcccaa 439
|||||
Db 351 AATAATCAACAATCAAGTTAAGAGCTTCAANAGGAGCAGCTCTCAAG:HTCICCCAA 420
QY 440 aagaacccctggactaagaagttactgctgctgctgctgctgctgctgctgctgctgctg 499
|||||
Db 421 AAAAGCACCTTGGTACTAAGAGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
QY 500 tgcgtgctaaagttcccaaaaaaatcacccgcgcgaagataaaaggtcccaaccca 559
|||||
Db 481 TGTGCTGCTAAAGTTCACGCAAAAAGATCACCGCGGAGTAAAGAG:TTTACGCCCA 540
QY 560 gaaggttcctgcccagaaagccacagccagaaagcagcctgctcccaaaagctcagaa 619
|||||
Db 541 GAAGGTTCCTGCCAGAAAGCCACAGCCAGAAAGCAGCGCCGCTCTCCAA:ATTCAGAA 600
QY 680 agtgcgaatcataaaagtataaa-gattctttttgacctatt 722
|||||
Db 561 AGTGGCANTATAAAAGTAAATNAAGGTTCTTTTACCGCTGTT 704

RESULT 5
LOCUS AU123120 728 bp mRNA linear: EST 23-OCT-2000
DEFINITION AU123120 NT2RM1 Homo sapiens cDNA clone NT2RM100150 5', mRNA
sequence.
ACCESSION AU123120
VERSION AU123120.1 GI:10947836
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ota.T., Nishikawa,T., Suzuki,Y., Kawai,Y., Ishii,S., Saito,K.,
Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T., Saito,Y., Kawai,Y.,
HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Kawai,Y.,
Ishii,S., Saito,K., Nakamura,Y., Nagai,T., Sugano,S., Isogai,T.)
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana-Kisarazu, Chiba 252-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5', 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. 728
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM1000750"
/clone_lib="NT2RM1"
/cell_type="teratocarcinoma"
/note="Vector: pUC19PL3; mRNA from undifferentiated NT2 neuronal
precursor cells"
220 a 173 c 181 g 153 t 1 other:-

BASE COUNT 220 a 173 c 181 g 153 t 1 other:-
ORIGIN

Query Match 94.3% Score 690.6 DB 5: Length 728:
```

--END--

```
Best Local Similarity 98.9%: Pred. No. 3.7e-134:
Matches 715: Conservative 0: Mismatches 5: Indels 3: Gaps 3:
QY 1 cgcctaacgctgcacaatggtttcagcgcttcgtggaggttgccgggtgctctatg 51
|||||
Db 5 CGCCTAACGCTGCACATGGTGTTCAGGCGCTTCGTGGAGGTGGCCGAGTGGCCTATG 65
QY 51 tctcctttggacctctatgcggaaaaattggctcgcgattgtagatgttattgataaaca 120
|||||
Db 66 TCTCCTTTGGACCTCATGCGGAAATTTGGTCGCGATTGTAGATGTATTGATCAGAACA 125
QY 121 ggcctttgctcaatgagcttcacccaadtgagagacagagccatgcctttcaaatgca 190
|||||
Db 126 GGCCTTTTGGTCGATGACCTTGCATCAAGTAGGAGACAGGCCATGCCCTTCAAGTGCA 195
QY 181 tgcagctcactgaatttcactctcctcaagtttctgcacagtgcccccacaaagatgtccgac 210
|||||
Db 186 TGCAGCTCACTGATTTCCTCTCAAGTTTCGCGACAGTGGCCACCAAGATATGTCGCAC 215
QY 241 aaccttggcagaaggcagacatcaatacaaaatgggcagccacacatggagccaaaaga 310
|||||
Db 245 AAGCCTGGCAGAGGAGAGCATCAATACAAATGGCGAGCACACGATGGGCCAAGAAGA 305
QY 301 ttaaacccagagaaagaaagccaaagatgacagattttgatcttttaaagttatgaag 350
|||||
Db 306 TTGAAGCCAGAGAAAGAAAGCCAAAGATGACAGATTTTGATCGTTTAAAGTTATGNAG 355
QY 351 caaagaaaatggagaaacagagataatcaagaatgaagttaagaagcttcaaaaggcagctc 420
|||||
Db 356 CAAAGAAAGTGGAGGAACAGAGATAATCAAGATGAAGTTAAGAGCTTCAAAAGGCGCTC 425
QY 421 tctgaaagcttctcccaaaaagccctgactaagggtactgctgctgctgctgctg 490
|||||
Db 426 TCTGAAAGCTTCTCCCAAAAAGCACCCTGCTACTAAGGGTACTGCTGCTGCTGCTG 495
QY 481 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 540
Db 486 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 545
QY 541 gtaaaagctccagccagaaaggttctctcccaagaagccacagccagaaagcagcgc 610
Db 545 GTAAAAGGCTCCAGCCAGAGAGGTTCTCTGCCAGAAAGCCACAGCGCCANAAAGCAGCGC 605
QY 601 ctgctccaaaagctcagaaggttcaaaaagctccagcccccagaaagcactgctccaaagg 650
Db 606 CTGCTCCAAAAGCTCAGAGGGTCAAAAAGCTCCAGCCCAAGAGCA-CGTCTCCAAAGG 654
QY 661 catctggcagaagaagcataag-tggcaatcat-aaaaagtaataaaggtctcttttgacc 718
|||||
Db 665 CATCTGCAGAAAGCATATAGTTGGCNAATCATANAAGTAATANAAGTTCTTTTGAAC 724
QY 719 ttt 722
Db 725 TGT 728

FEATURES
7
AL531631
LOCUS AL531631
DEFINITION AL531631 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSQDM002YH21 5
prime, mRNA sequence.
ACCESSION AL531631
VERSION AL531631.1 GI:12795124
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 818)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
```


Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Location/Qualifiers
1. .818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DM002YH21"
/clone_lib="L11_NFL001_HBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 5; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 5 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 510 8371 Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com"

BASE COUNT 271 a 172 c 199 g 175 t 1 others
ORIGIN

Query Match 94.2%; Score 680.4; DB 9; Length 818;
Best Local Similarity 99.0%; Pred. No. 4.1e-134;
Matches 697; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY 19 ttgtgttcaggcgttcgtggaggttgccgggtgagccctatctctcttttgacacctatg 78
|||||
DB 1 TGTGTTCAGGCGCTTCGTGGAGGTGGCGGGTGGCCATGTCCTTTTGACCTCATG 50
|||||

QY 79 ccggaataatgttcgcgatttagatgttatgatcagacagggcctttgctcatgac 138
|||||
DB 61 CCGGAAATATGTCGCGATTGTAGATGTTATTCATCAGAACAGGCGCTTGGTCGATGGAC 120
|||||

QY 139 cttgcactcaagtgaagacacagggccatcctttcaagtgcagctcactgatctca 198
|||||
DB 121 CTTCACCTCAAGTAGAGGACAGGCCATGCCCTTCAAGTGCATCAGCTCATGATCA 180
|||||

QY 199 tctcaagtttctgacatgcccaccagaagtatgtccgacaagccttggcagaagcag 258
|||||
DB 181 TCTCAAGTTTCGCGACAGTGCACACAGCAAGATATGTCGACAGCCCTGGCAGAGGCGAG 240
|||||

QY 259 acatcaatacaaaatggcagccacacgatggcccaaaagatgaagccagaagaaga 318
|||||
DB 241 ACATCAATACAAAATGGGAGCCACACAGATGGCCNAGAGATTTGAAGCCAGAGAAAGGA 300
|||||

QY 319 aagccagatgacagattttgatcgttttaaaagttatgaagcaagaagaaatgagaa 378
|||||
DB 301 AAGCCAAGATGACAGATTTTGATCGTTTAAAGTTATGAAGCAAGAAATGAGCAACA 360
|||||

QY 379 gaataatcaagaatgaagtaagaagcttcaaaagcagactctctgaaagcttctcca 438
|||||
DB 361 GAATAATCAAGAAATGAAGTTAGAAGCTTCAAAAGCGAGCTCTCTGAAAGCTTCTCCA 420
|||||

QY 439 aaaaacacctgtactaaggtaactgctgctgctgctgctgctgctgctgctgctg 498
|||||
DB 421 AAAAAGCACCCTGGTACTAAGGTA-----CTGCTGCTGCTGCTGCTGCTGCTGCTG 474
|||||

QY 499 ctgctgctgctaaagttccagcaaaaaagatcacccgcgcgcgagtaaaaaaggtctcagcc 558
|||||
DB 475 CTGCTGCTGCTAAAGTTCACGCAAAAAGATCACCCGCCGCGAGTAAAAGGCTCCAGGCC 534
|||||

QY 559 agaaggttctctcccaagaagccagagccagaagaagcagcgcctgctctccaaaagctcaga 518
|||||
DB 535 AGAAGGTTCTTCGCCAAGAAAGCCAGAGCCAGAAAAGCAGCGCTTCTCCAAAAGCTCAGA 594
|||||

QY 619 aggtgtcaaaagctccagccctcagaagcacctgcttcccaagagcatctaggcaagaagcat 678
|||||
DB 595 AGGGTCAAAAAGCTCCAGCCCGAGAAAGACCTTCTCCAAAGGCGATCTGGCAAGAAAGCAT 654
|||||

QY 579 aagtgccaatcataaaagaataaagtgcttcttttgaccttt 722
|||||
DB 655 AAGTGGCAATCAATAAAAGTAAATAAAGTTCTTTTGTACCTGTT 698
|||||

RESULT 8

BM472466 723 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6464185 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5533444
DEFINITION 5', mRNA sequence.
ACCESSION BM472466
VERSION BM472466.1 GI:18521508
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 723)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DIP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1231 row: f column: 13
High quality sequence stop: 637.
Location/Qualifiers
1. .723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5538444"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

FEATURES
source

BASE COUNT 235 a 167 c 178 g 143 t
ORIGIN

Query Match 94.2%; Score 679.8; DB 10; Length 723;
Best Local Similarity 98.0%; Pred. No. 5.5e-134;
Matches 703; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 5 aacgtcccaacatagtattcaggcgttcctggaaggttgccgggtgacctatctcc 65
|||||
DB 1 AACCGCGCAACATGGTGTTCAGGCGCTTCGTGGAGGTGGCGGGTGGCCCTATGTCCTCC 50
|||||

QY 55 ttggacacctcatgccggaaaattgctgcgattgtagatgttattgatcagacagggct 125
|||||

DB 51 TTGGACCTCATGCCGGAATTTGGTCGGATTGTAGATTGTTATTGATCAGAACAGGCT 120
|||||

QY 125 ttggtcgatggaccttgacctcaagtggaggagacagggccatgctcttcaagtgcacag 185
|||||

DB 121 TTGTGCGATGGACCTTGCACCTCAAGTGAGGAGACAGGCGCATGCCCTTTCAGGTGCATGCAG 180
|||||

QY 185 ctccactgatttcactcctcaagtttctgcacagtgccaccagagaagtatgtccgacaagcc 245
|||||

DB 181 CTCACCTGATTTCATCCTCAAGTTTCCGACAGTGCAGCCACCAAGATATGTCGCAAGCC 240
|||||

QY 245 tggcagaaggcagacatcaatacaaaatgggcagccacacagatggcccaagaagattgaa 305
|||||

DB 241 TGGCAGAGGCGAGACATCAATACAAAATGGGCGGCGCCACGACGATGGCCCAAGAAATTGNA 300
|||||

```

Oy 305 gccagagaaagaaagccagagattttgactgttttaaaagttaagagagaaag 365
|||||
Db 301 GCCAGAGAAAGAAAGCCAGATGACAGATTTTGAATGTTTTAAAGTTATGAGAGCAAG 360
Oy 365 aaataggagaaagaaatcaagaatgaagttaagaagtttcaaaagggcagctctcttg 425
|||||
Db 361 AAAATGAGGAACAGATAATCAAGATGAATTAAGAGGTTTCAAAAGGCAAGTCTCTG 420
Oy 425 aaagcttctcccaaaagacacctgtactaaggggtactgtctgtctgtgtgtgt 485
|||||
Db 421 AAAGCTTCTCCCAAAAGACCTGTGTTACTTAAGGGTA-----CTGTGTGTCT 468
Oy 485 gctgtctactgtctgtctgactaaagttccagcaaaaagaaagacccgcgcgcgaataaa 545
|||||
Db 469 GCTGTCTGTCTGTCTGTCTGTCTGAATTCAGCAAAAGATCACCGCGTCCAGTAA 528
Oy 545 aaggtctccagccagagaggttctctgcccaaaagccacagggccagaaagcaagctgtgt 605
|||||
Db 529 AAGGCTCCAGCCAGAGGTTCTGCCAGAAAGCCACAGCCAGAAAGCAAGGCTGTCT 588
Oy 605 ccaaaagctcagaagggtaaaagctccagccagaaagcactgtctccaaagggatct 665
|||||
Db 589 CCAAAAGCTCAGAGGGTCAAAAGCTCCAGGTCAGGTCAGAAAGCAAGCTGTCTCCAAAGCAATCT 648
Oy 665 ggaagaaagacataagtggaatcataaaagtaataaaagttcttcttttgaacttt 722
|||||
Db 649 GGCAGAAAGCATAGTGGCAATCATAAAAGTAATAAGGTTCTTTTGAATGTTGTT 705

RESULT 9
BG395368 730 bp mRNA linear EST 12-MAR-2001
LOCUS 502459230F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4581479 5'
DEFINITION mRNA sequence.
ACCESSION BG395368
VERSION BG395368.1 GI:13289914
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
AUTHORS Mammalia: Euthera: Primates: Catarrhini: Hominiidae: Homo.
1 (bases 1 to 730)
TITL NIH-MGC http://imgc.nci.nih.gov/
JOURN National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs@mail.nih.gov
Tissue Procurement: AICC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: http://imgc.llnl.gov
Plate: L12W1302 row: d column: 24
High quality sequence stop: 727.
Location/Qualifiers
1..730
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4581479"
/clone_lib="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; vector: pCIB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by: oligo-dT priming. directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACAGAG(G). Library constructed by Ling Hong
in the Laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
231..a 170..c 183..g 145..t

```

```

Oy 12 gccaaatagtggttcgaagccttcgtgagagttggccgggtggcctatgtctcttggga 71
|||||
Db 1 GCCAAATGTTGTTTCAGGCGCTTCGTGGAGGTTGGCCGGGTGCTATGTCTCTTTGGA 60
Oy 72 cctcatccgagaaatggtcgcgattgtagatttattgatcagaacagggcttggctc 131
|||||
Db 61 CTTATGCCGGAATTTGGTCCGGAATTTAGATGTTATTGATCAGAACAGGGCTTTGGTC 120
Oy 132 gatggacttgcactcaagtgaagagacagggccttctcaagtcatgcaagctcact 191
|||||
Db 121 GATGGACCTTGCACATCAATGAGGAGACAGGCCATGCTTTTCAAGTGCATGCAGCTCACT 180
Oy 192 gatttcactcgaagttctgcacagtgccaccagaaagtatgtccaaagcctggcag 251
|||||
Db 181 GATTTTCATCTCAAGTTTCCGCACAGTCCACAGAGTATGTCCGACAAAGCTTGGCAG 240
Oy 252 aagcgacatcatcaatacaaaatggcgagccacacgatggccaaagaaagattgaagccaga 311
|||||
Db 241 AAGGCACACATCAATACAAATGGCGACCCACAGATGGGCCAGAAAGATTGAAGCCAGA 300
Oy 312 gaaaggaagccaaagatgcagattttgatcgtttttaaagtattgaaggcgaagaaatg 371
|||||
Db 301 GAAAGGAAGCCAGATGACAGATTTTGATCGTTTTTAAAGTTATGAAGGCAAGAAAGATG 360
Oy 372 agaaacagaataatcaagaatgaagtttaagaagcttcaaaagcagctctcctaaagct 431
|||||
Db 361 AGGAACAGAAATATCAAGAATGAAGTTAAGAAGCTTCAAAAGGCGCTCTCTTGAAAGCT 420
Oy 432 tctcccaaaaagccactgtactaaggttactgtctgtctgtctgtctgtctgtctgtct 491
|||||
Db 421 ICTCCCAAAAAGCACCTGTGTACTAGGGTACTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 480
Oy 492 gctgtctctctctctcctaaagttccagcaaaaagaaatcacccgcgcgcgagataaaagct 551
|||||
Db 481 GCTGTCTGTCTGTCTGTCTAAGTTCAGCAAAAAGAAATACCCGCGCGAGTAAAGAGCT 540
Oy 552 ccagccagaaagttctctgccagaaagccacagggccagaaagcagcgcctgtctccaaa 611
|||||
Db 541 CCAGCCCAAGAGGTTCTGCCCAAGAAAGCCACAGGCCGAGAAAGAGCGG-CTGTCCAAA 599
Oy 612 gctcagaaggggtcaaaaagctccagcccaagaaagcactgtctccaaagcagctatggcag 671
|||||
Db 602 GCTCAGAGGGTCAAAAGGCTCCAGCCAGAAAGACCTGTCTCCAGAGGCAATCTGGCAAG 659
Oy 672 aaagcataagtcgaatcataaaagtataaaggttcttttttaaccttt 722
|||||
Db 650 AAAGCATAAGTGGG-ATCATAAAAGTAATAGAGGTTCTTTTCGACCTGGT 709

RESULT 10
BG1914072 746 bp mRNA linear EST 16-CTT-2001
LOCUS 503182003F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246184 5'
DEFINITION mRNA sequence.
ACCESSION BG1914072
VERSION BG1914072.1 GI:16178382
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
AUTHORS Mammalia: Euthera: Primates: Catarrhini: Hominiidae: Homo.
1 (bases 1 to 746)
TITL NIH-MGC http://imgc.nci.nih.gov/
JOURN National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs@mail.nih.gov

```

Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM11621 row: e column: 01
 High quality sequence stop: 713.
 Location/Qualifiers

FEATURES
 source

1..746
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5246184"
 /clone_lib="NIH_MGC_121"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 21 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb. Insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."

BASE COUNT 235 a 175 c 184 g 151 t
 ORIGIN

Query Match 92.8%; Score 677.4; DB 10; Length 746;
 Best Local Similarity 99.0%; Pred. No. 1.8e-133;
 Matches 692; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Oy 17 catgtgttcacagccttcgtgagagcttgcgggtgacctatctctcttgcacctca 75
 Db 1 CAATGTGTCAGCGCTTCGTGGAGGTTGGCGGGTGGCTATGCTCTCTTGGACCTCA 60
 Oy 77 tcccgaaaattgctgcgattgtatattatcagaacagggctttggtcgatcg 135
 Db 61 TCCCGAAAATTGGTCGGATTGTAGATGTTATGATCAGACAGGGCTTTGTCGATG 120
 Oy 137 accttgcactcaagttagagagacaggccatgcctttcaagtgcactgcactgattt 195
 Db 121 ACCTTGCACTCAAGTGAAGACAGCGCCATGCTTTCAAGTCATGACGACATGATT 180
 Oy 197 catctcaagatttcgcacagtgcacacagaaagtatctgcacaaagctgcacaaaggc 256
 Db 181 CATCTCAAGTTTCGCACAGTCCGACACAGAAATATGCCACAAGGCTGGCAAGG 240
 Oy 257 agacatcaatacaaaatggcagccacacacgattgggccaagaattgaagccagagaag 316
 Db 241 ACACATCAATACAAATGGCGACCCACACGATGGGCCAAGAGATTGAAGCCAGAG 300
 Oy 317 gaaagccaaagtacagattttgatcttttaagtattgaaggcagaagaaatagaaa 376
 Db 301 GAAAGCCAAAGATGACAGATTTTGTATCGTTTAAAGTTATGAAGGCAAAAGAAATGAGGAA 360
 Oy 377 cagaataatcaagaatgaagtttaagaacttcaaaagcagactctctgaagactctcc 436
 Db 361 CAGAATAATCAAGAATGAAGTTTAAAGCTTCAAAAGGCGAGCTCTCTCAAGCTCTCC 420
 Oy 437 caaaaaagcacttgatctactaagggttactgctctgctgctgctgctgctgctgctgct 496
 Db 421 CAAAAAGCACCCTGTTACTAAGGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 Oy 497 tctctctctctaaagttccagcaaaaaaatcatccgctcgagtagtaaaaggtccagcc 556
 Db 481 TCTCTCTCTGCTAAAGTTCCAGCAAAAAAGATCATCCGCGCGAGTAAAGGCTCTCAAG 540
 Oy 557 ccagaagttctctcccaagaagccacagagccagaaagcagcctgctctcaaaagctca 616
 Db 541 CCAGAAGGTTCTTCCCAAGAAAGCCACAGGCCAGAGAGAGCGGCTGTCTCAAGAGCTCA 600

Oy 617 caaggttcaaaaagct-ccagcccaagaagcaccctgctctccaaagcactctgacagaaaag 675
 Db 601 GAAGGTCAAAAGGTCCTCCAGCCAGAACGACCTGCTCTCAAGGCACTCTGCAAGAAAG 660
 Oy 675 cataagtgacatcataaaaagtaataaaggttcttttt 714
 Db 661 CATAGTGGCAATCATACCAAGTATATAAGGTTCTTTT 699

RESULT 11
 BG260421 732 bp mRNA linear EST 13-FEB-2001
 LOCUS 602371552F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4479420 5'
 DEFINITION mRNA sequence.
 ACCESSION BG260421 GI:12770237
 VERSION BG260421.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 732)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rgs@bbs.rmail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10311 row: h column: 13
 High quality sequence stop: 718.

FEATURES
 source

1..732
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4479420"
 /clone_lib="NIH_MGC_93"
 /tissue_type="transitional cell papilloma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 225 a 175 c 185 g 147 t
 ORIGIN

Query Match 93.5%; Score 675.2; DB 10; Length 732;
 Best Local Similarity 98.6%; Pred. No. 5.1e-133;
 Matches 702; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

Oy 2 gcctaacactgcaacatggttctcagcgtcttcgaaggttgccgggtgacctatct 61
 Db 1 GCCTAACCTGCAACATGGTGTTCAGCGCTTCGTGGAGGTGGCGGGTGGCCTAIGT 60
 Oy 62 ctctcttgagacctatgcgcaaaaattggtcgagattgtagattgattgatcagaacag 121
 Db 61 CTCCTTTGGACCTCATSCGGNAATTTGTCGCGATTGTAGATTGTTATTGATCAGAACAG 120
 Oy 122 ggccttgatgatgacaccttgacctcaagttaggagacaggccatgccttcaagtgcatt 181
 Db 121 GGCCTTTGGTTCGATGGACCTTGACCTCAAGTGAAGAGACAGGCCATGCCCTTCAAGTGCA 180
 Oy 182 ccagctcactaatctcactcaagtcttcacagatccaccacagaagtagtctccgaca 241
 Db 181 GCAGCTCACTGATTTTCACTTCCTCAAGTTTCCGCAAGTTCGCCACCAAGAGTATGTCGACA 240

242 agctgacgagagagacacatacaataaaatggggcagccacacagatgggccaagaagat 301
 Db AGCTGGCAGAGGACACATCAATACAAAATGGCAGCCACACAGATGGGCAAGAAGAT 300
 Oy tgaagccagagaaaggaagccaaagatgacagattttgactgttttaagttatgaagcc 361
 Db TGAAGCCAGAGAAAGGAAAGCCCAAGATGACAGATTTTGTATCGTTTAAAGTTATGAAGCC 360
 Oy aagagaaatagggaacagaataatcaagaatgaagtttaagaagcttcaaaagggcagctct 421
 Db AAAGAAAATAGGAATCAGATAATCAAGATGAAGTTAAAGAGCTTCAAAAGGACAGTCT 420
 Oy cctaaagcttctcccaaaagacactgtactaaaggtactctctgtctgtctgtc 481
 Db CCTGAAGCTTCTCCCAAAAAGCACCCTGGTACTTAAGGGTACTGCTCTGCTGCTG 480
 Oy tctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 541
 Db TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
 Oy taagaagcttccagccagagaggttctgtccgagaaagccagagccagagagagagcc 601
 Db TAAAGAGCTTCCAGCCAGAGAGGTTCTGCTCCAGAAAGCCACAGGCGCAGAGCAGGCGC 600
 Oy tctcccaaaagctcagaagggctc-aaaaagctccagcccaaaagagcagcagcagcagc 660
 Db TGTCTCCAAAAGCTCAGAAAGGTCNAAAAGCTCCAGCCCAAGAGCAAGTCTCCAGAGG 660
 Oy catctgcagaaagcagtagtgccatcataaaagtaataaaggtctttt 712
 Db CATCTGCAAGAAAGCAATAAGTGGC-ATCATAAGAGTACTAAGGTCTCTTT 711

RESULT 12

BM043710 503620502F1 NIH_MGC_40 705 bp mRNA linear EST 07-NOV-2001
 LOCUS 503620502F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5446200 5'
 DEFINITION mRNA sequence.
 ACCESSION BM043710
 VERSION BM043710.1 GI:16772966
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: rstraubs@mail.nih.gov
 Tissue Procurement: DCID/DIP
 CDNA Library Preparation: Ling Hong/Pubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov/
 Plate: LLCM1928 row: c column: 01
 High quality sequence stop: 704.
 Location/Qualifiers
 1..705
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5446200"
 /clone_lib="NIH_MGC_40"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: prostate; Vector: pOIE"; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed by

FEATURES

Source

RESULT 13
 AL528184
 LOCUS AL528184 LTI_NFL003_NBC3 712 bp mRNA linear EST 13-FEB-2001
 DEFINITION AL528184 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS00C025YL02 5'
 prime, mRNA sequence.
 ACCESSION AL528184
 VERSION AL528184.1 GI:12791677
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-CDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: This is a NIH_MGC Library.

BASE COUNT 215 a 165 c 182 g 142 t
 ORIGIN
 Query Match 93.4%; Score 674.4; DB 10; Length 705;
 Best local Similarity 98.7%; Pred. No. 7.6e-133;
 Matches 591; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

Oy 25 caggccgtctcgtgaggttgccgggtggcctatgtctctctttggaccctcatcccgaaa 85
 Db 2 CAGCGCCTTCGTGAGGTGGCCGGGTGGCCTATGTCCTTTGGACCTCATCCGGAA 61
 Oy attgctgcgagattgtaagtattatcatcagaacagagcttggtagagaccttcac 145
 Db 62 ATGTGTCGCGATTGTAGATTATGATCAGAACAGGGCTTTGTCGATGGACCTTGGAC 121
 Oy tcaagttagagagacagagccatgccttttcaagtgcacgtcagctcactgatttcctcaa 205
 Db 122 TCAAGTAGGAGAGACAGGCCATGCTTTCAAGTGCATGCAGCTCACATTTCATCTCAA 181
 Oy attcttcacagctgccacacagaagtatgtccaaagcctggcagaagcagacatcaa 265
 Db 182 GTTTCOCACAGTGGCCACAGAAAGTATGTCGACAAAGCCTGGCAGAAAGCAGACATCAA 241
 Oy tacaataatggcagccacacacatgggccaaagaattgaagccagagaagaagcccaa 325
 Db 242 TACAAAATGGCAGCCACACAGATGGGCCAAGAAGATTGAAGCCAGAGAAAGGAAAGCCAA 301
 Oy gatcagagattttgatcgttttaagttatgaaggcaaaagaaatagaacagataat 385
 Db 302 GATGACAGATTITGATCGTTTTAAAGTTATGAAGGCAAGAAATGAGGAACAGAAATAT 361
 Oy caagaatgaagttaagaagcttcaaaagcagctctctctgaaagcttctcccaaaaagc 445
 Db 362 CAAGAATGAAGTTAAGAAGCTTCAAAAGCGAGCTCTCTCTGAAAGCTTCTCCCAAAAAGCG 421
 Oy acctgtactaaggdta---ctgctactctgctgctgctgctgctgctgctgctgctg 502
 Db 422 ACCTGTACTAAGGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481
 Oy tgcgtctaaagttccagcaaaaaaatcaccgcgcgagataaaagagctcccaagccagaa 562
 Db 482 TCGTGCTAAAGTTCACGACAAAAGATCACGCCGCGAGTAAAGAGCTCCAGCCAGAA 541
 Oy ggttcctgccagaaaagccacagccagaaagcagagcctgtccaaaagctcagaagg 622
 Db 542 GGTTCCTGCCAGAAAGCCACAGCCAGAAAGCAGGCCCTGTCTCCAAAAGCTCAGAGGG 601
 Oy tcaaaagctccagcccaaaagcaccctgtctccaaagcattctggcagaagaagcataagt 682
 Db 602 TCAAGAAGCTCCAGCCCAAGAGCAGCTCTCCAAAGGATCTGGCAAGACAGCATAAGT 661
 Oy ggcacatcaaaagtaataaaggttcttttgacctgtt 722
 Db 662 GSCAATCAATAAAGTAATAAAGGTTCTTGTAGAGCTGTT 701

```
REFERENCE
AUTHORS 1 (bases 1 to 712)
TITLE Li, W.B., Gruber, C., Jesse, J., and Polayres, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91005 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 712
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0PC025YL02"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: Brain; Vector: pCMVSPORT 5; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prim' end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 5
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 510
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 227 a 161 c 177 g 145 t 2 others
ORIGIN
source

Query Match 93.3%; Score 673.4; DB 9; Length 712;
Best Local Similarity 98.3%; Pred. No. 1.2e-132;
Matches 593; Conservative 1; Mismatches 12; Indels 9; Gaps 1;

Oy 18 atggttcaggcgttcgtgaggttgccgggtgagcctatctcccttgacctcat 77
Db 1 ATGGTGTTACGGCGCTTCGTGGAGTTGGCCGGTGGCTATGTCCTTTGGACCTCAT 50

Oy 78 gccggaattgtgcgatttagattgtattgatacagaacaggcctttgctcatgga 137
Db 61 GCCGGAAATTTGTCGCGATTGTAGATTATTTGATCAGAACAGGCGTTTGGTCGATGCA 120

Oy 138 ctttgactcaatgaggagacagggcctgcttccaaatgcatcagctcactaatttc 197
Db 121 CTTTGACCTCAAGTAGGAGACAGGCGCATGCCCTTTCAGTGATGATGATGCTACTGATTTTC 180

Oy 198 atcctaagtttctcacagtgatccaccacagaagatgtgtccgacaagcctggcagaaggca 257
Db 181 ATCCCTCAAGTTTCCGCACAGTGCACCAAGATATGTTCGACAAAGCCCTGGCAGAAAGGCA 240

Oy 258 gacatcaatacaaaaatggcgaccacacacatggcccaaaaatgaaacccagaaaagg 317
Db 241 GACATCAATACAAAATGGCAGCCACACGATGGCCCAAGAAATTTGAAGCCAGAAAGG 300

Oy 318 aaagccaagatacagattttatcgttttaagttatgaagcagaagaaaatgaggaac 377
Db 301 AAAGCCCAAGATGACAGATTTTATGCTGTTTAAAGTTATGAAGCAAGAAAGAAATGAGGAAC 360

Oy 378 agataatcaagaataaagttaagaagcttcaaaagcagctctctctaaagcttctccc 437
Db 361 AGAATATCAAGAAATGAAGTTAAGAGCTTCAAAAGGCGAGCTCTCTGAAAGCTTCTCCC 420

Oy 438 aaaaagcacccttgactaagggtagctgctgctgctgctgctgctgctgctgctgct 497
Db 421 AAAAAAGACCTGGTACTAAGGGTA-----CTGCTGCTGCTGCTGCTGCTGCTGCT 471

Oy 498 gctgctgctcctaaagttccacaaaagatcacccgcccagagtaaaaaggctccagcc 557
Db 472 GCTGTGCTGCTTAAAGTTCACGCAAAAAGATCACCGCGCGAGTTAAAGAGCTCCAGCC 531

Oy 558 cagaaggttctctcccaagaaacacagccagagccagaagcagcctgctcccaaaagctcag 617
Db 558 cagaaggttctctcccaagaaacacagcagtagtgcgcagaaagctgctccgagaagg 255

Db 532 CAGAAAGTTCTCTCCAGAAAGCCACAGGCCAGAAAGCAGCGCTCTCTCAAAAGCTCAG 591
Oy 618 aaggggtcaaaaagctccagccagcaaacgacctgctcccaagcattctgcaagaagca 677
Db 592 AAGGGTCAAAAGCTCCAGCCAGAAAGCAGCTCTCTCAAAAGCAGCTCTGCAAGAAAGCA 651
Oy 578 taagtgcgaatcataaaaagataaataaagttctttttgacctgatt 722
Db 552 TAAGTGGCAATCAATAAAGATAAAGGTTCTTTTGGACCTGTT 696

RESULI 14
LOCUS BM480399
DEFINITION AGENCOURT_6458573 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5574262
5', mRNA sequence.
ACCESSION BM480399
VERSION BM480399.1 GI:18529441
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1116)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@pshs-research.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1232 row: j column: 23
High quality sequence start: 27
High quality sequence stop: 636.
Location/Qualifiers
1. 1116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5574262"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: small intestine; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH-MGC Library."
BASE COUNT 298 a 308 c 254 g 211 t 45 others
ORIGIN
source

Query Match 93.0%; Score 671.4; DB 10; Length 1116;
Best Local Similarity 98.2%; Pred. No. 3.5e-132;
Matches 594; Conservative 0; Mismatches 1; Indels 12; Gaps 1;

Oy 15 acatggtgttcaggcgttcgtgaggttgccgggtgagcctatgctcctttgacctc 75
Db 38 ACATGGTGTTCAGGCGCTTCGTGGAGTTGGCCGGTGGCTATGCTCTCTTTGGACCTC 97

Oy 75 atgccgaaaaattggtcgcatttagattgattgatacagaacaggccttggctcag 135
Db 98 ATGCCGAAAATTTGTCGCGATTGTAGATGTTATTGATCAGAACAGGCTTTGGTCGATG 157

Oy 175 gaccttgacctcaagtgaggagacagggcctatgcctttcaaatgcatgcagctcactgatt 195
Db 158 GACCTTGACACTCAAGTAGGAGACAGGCCATGCCCTTTCAAGTGCATGCAGCTCACTGATT 217

Oy 195 tcatcctcaagtctctcacagtagtccccaccagaagtagtgcgcagaaagcctgaggaagg 255
```

```

|||||
128 TCATCTCAAGTTCGGACAGTGGCCACCAAGATATGTCGAAAGCCGGCAGAAAG 277
256 cagacatcaatacaaaatgggcagccacacagatggcccaagaagattgaagccagagaaa 315
278 CAGACATCAATACAAATGGGAGCCACAGGATGGCCCAAGAAGATTGAAGCCAGAGAA 337
316 gaaagcccaagatgaagaattttgattgattttaaagttatgaagcgaagaaatgagga 375
339 GGAAGGCCAAGAATGACAGATTGATCGTTTAAAGTTATGAAGGCAAGAAATGAGGA 397
375 ecagaataatcaagaataaagttgaagcttcaaaagcagctctctgaaagctttctc 435
398 ACAGAAATCAAGATCAAGTATAGATGAGCTTCAAAAGGCAAGCTCTCTGAAAGCTTCTC 457
436 ccaaaaaagcaccctgtactaagggtactgtgtgtgtgtgtgtgtgtgtgtgtgtgt 495
458 CCAAAAAGGCAAGCTTGTACTAAGGTA-----CTGCTGCTGCTGCTGCTGCTG 505
495 ctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 555
505 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 565
555 cccagaaggttctgtccagagaagccacagccagagaagcagcagcagcagcagcagc 615
565 CCCAGAAGGTTCTTCCAGAGAGCCACAGGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 625
615 gaaaggttcaaaagctccagccagagaagcagcagcagcagcagcagcagcagcagc 675
625 AGAAGGGTCAAAAGCTCCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 685
676 cataagtgcccaatcetaaaagtaataaagattcttttaaccctgt 722
686 CATAGTGGCAATCATAAAGATTAATGAAGGTTCTTTTGAACCTGTT 732

```

RESULT 15

```

LOCUS      BG437749          724 bp    mRNA    linear,   EST 14-MAR-2001
DEFINITION 60248926F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4521783 5',
            mRNA sequence.
ACCESSION  BG437749
VERSION    BG437749.1 GI:13344255
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1. (bases 1 to 724)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: rgs@fmail.nih.gov
Tissue Procurement: DCID/DIF/Gazdar
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1385 row: d column: 08
High quality sequence stop: 704.
Location/Qualifiers
1..724
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4521783"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pCMV7; Site: 1; XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned

```

FEATURES

SOURCE

Search completed: June 21, 2002, 13:49:45
Job time: 6885 sec

into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library."

BASE COUNT 229 a 165 c 185 g 144 t
ORIGIN

Query Match 92.8%; Score 670; DB 10; Length 724;
Best Local Similarity 98.3%; Pred. No. 6.5e-132;
Matches 700; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

```

Oy 12 gccaacatggttcagagccttcgtgagaggttgccggtggcctatgtctcttggga 71
Db 1 GCCAACATGTTGTCAGGCGCTTCGTGAGGTTGGCCGGTGGCTATGTCTCTTTGGA 60
Oy 72 cctcatgccggaataatggttcgcatgtatgtattgatcagaaacagagctttggtc 131
Db 51 CCTCATGCCGGAATAATGGTCGCAITGTAGATGTTATTGATCAGAACAGGCTTTGGTC 120
Oy 132 gatgaccttcacacaaagtgaagagagagcagcagctgtttcaagtcacatgcagctcaat 191
Db 121 GATGACCTTTCACACTCAAGTGAAGAGAGACAGCCATGCTTTCAAAGTCATCGACTCACT 180
Oy 192 gatttcacctcaagtttctcacagtcgcccacagagatgtcccaacagcctggcag 251
Db 181 GATTTCATCTCAAGTTCCGACAGTGGCCACCAAGATGTATGTCCGCAAGCCTGGCAG 240
Oy 252 aaggcagacatcaatacaaaatgggcagccacacagatggcccaagaagattgaagccaga 311
Db 241 AAGCAGACATCAATACAAATGGCAGCCACACAGATGGCGGCAAGAAAGATTGAAGCCAGA 300
Oy 312 gaaaggaagcccaagatcacagattttgatcgttttaagttatgaaggcaagaaatg 371
Db 301 GAAAGGAAGCCAAAGATGACAGATTITGTCGTTTTAAAGTTATGAAGGCCAAAGAAATG 350
Oy 372 agaacagaataatcaagaatgaagttaagaagcttcaaaagcagcgtctcctgaaagct 431
Db 361 AGGAACAGATAATCAAGAATGAAGTTAGAGCTTCANAAAGCAGCTCTCTCTGAAAGCT 420
Oy 432 tctcccaaaaaagcaccctgactaagggttactgtctgtctgtctgtctgtctgtct 491
Db 421 TCTCCCAAAAAGCACCCTGTACTAAGGTA-----CTGCTGCTGCTGCTGCTGCT 474
Oy 492 gctgtgtctgtctgtctaaagttccagcaaaaaagatcaccgcccgagtaaaagcct 551
Db 475 GCTGCTGCTGCTGCTAAGTTCAGCAAAAAAAGATCACCOCGCCGAGTAAAGGGCT 534
Oy 552 ccagcccaagaagttcctgccagaagccacagagccagaagcagcctgtctccaaaa 611
Db 535 CCAGCCCAAGAAGTTCTGTGCCAAGAACCCACAGGCCAGAAAGCAGCGCTGCTCCAAA 594
Oy 612 gctcagaaggtcaaaaactccagcccaagaagacacctgtctccaaagcagctgcaag 671
Db 595 GCTCAGAAGGTTCAAAAGCTCCAGCCCAAGAAAGCAGCTGCTCCAAAGCATCTGCAAG 654
Oy 672 aaagca-taagtgagcgaatcataaaagtaataaagttcttttgacctgtt 722
Db 655 AAAGCATTAAAGTGGCAATCATAGAAGTAAATGAAGTTCTTGTGAGCTGTT 705

```

